

XX Claim 8; Page 702; 747pp; English.
 PS
 XX The present invention relates to novel plant transcription factors from
 CC Eucalyptus grandis or Pinus radiata. The present sequence is one such
 CC transcription factor. The transcription factor may be used to produce a
 CC plant having modified gene expression such as a woody plant e.g. a
 CC eucalyptus, pine, acacia, poplar, sweetgum, teak, or mahogany species or
 CC to modify the activity of a polypeptide in a plant. The transcription
 CC factors of the present invention are members from the following families
 CC of regulatory proteins: bzip, bzip family of G-box binding factors, basic
 CC helix-loop-helix zipper, homeotic/homeodomain/homeobox/MADS, homeodomain
 CC zipper, LIM domain, AP2 and EREBs, zinc finger domains of type 2
 CC Cys2His2, CCAT box elements and MYB.
 XX
 SQ Sequence 96 AA;

Query Match 100.0%; Score 24; DB 21; Length 96;
 Best local Similarity 100.0%; Pred. No. 4,7e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSV 5
 |||||
 Db 20 VSGSV 24

Search completed: September 6, 2001, 16:43:30
 Job time: 359 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:49:51 : Search time 231.42 Seconds
(without alignments)
4.574 Million cell updates/sec

Title: US-09-603-713-23
Perfect score: 42
Sequence: 1 XGERGFY 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SPREMBL_16:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_phase:*
11: sp_plant:*
12: sp_rhodent:*
13: sp_unclassified:*
14: sp_vertebrate:*
15: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	90.5	66	13	091163
2	38	90.5	106	13	091807
3	37	88.1	1259	10	022326
4	37	88.1	1259	10	023102
5	37	88.1	1260	10	09RHN8
6	37	88.1	1261	10	039130
7	35	83.3	141	2	046795
8	35	83.3	188	2	09RDX9
9	35	83.3	349	2	09PCX4
10	35	83.3	551	10	043607
11	35	83.3	2586	5	P91871
12	34	81.0	116	13	091161
13	34	81.0	117	13	091476
14	34	81.0	145	13	091475
15	34	81.0	149	13	091231
16	34	81.0	155	13	091162
17	34	81.0	159	13	093607
18	34	81.0	161	13	091230
19	34	81.0	162	13	073720

20	34	81.0	182	13	042289	042289 oreochromis
21	34	81.0	182	13	P79824	P79824 oreochromis
22	34	81.0	184	13	042336	042336 myoxocephal
23	34	81.0	185	13	057436	057436 paralichthy
24	34	81.0	185	13	09Y157	09Y157 acanthopagr
25	34	81.0	186	13	093527	093527 paralichthy
26	34	81.0	186	13	09PSX5	09PSX5 paralichthy
27	34	81.0	188	13	P81268	P81268 oncorhynch
28	34	81.0	188	13	091965	091965 oncorhynch
29	34	81.0	340	10	09FLP7	09FLP7 arbidopsis
30	34	81.0	343	1	09H190	09H190 thermoplas
31	34	81.0	1518	5	09NCP5	09NCP5 dictyostell
32	33	78.6	559	2	09K0S8	09K0S8 neisseria m
33	33	78.6	580	2	09JY22	09JY22 neisseria m
34	33	78.6	580	2	09J001	09J001 neisseria m
35	33	78.6	580	2	09J0J5	09J0J5 neisseria m
36	32	76.2	273	5	026548	026548 schistosoma
37	32	76.2	292	2	09U0U1	09U0U1 neisseria m
38	32	76.2	331	5	09U2S6	09U2S6 caenorhabd1
39	32	76.2	370	2	053159	053159 mycobacteri
40	32	76.2	374	2	068209	068209 escherichia
41	32	76.2	374	2	068269	068269 salmonella
42	32	76.2	374	2	09ZIS0	09ZIS0 escherichia
43	32	76.2	374	2	09R2L8	09R2L8 escherichia
44	32	76.2	447	5	09U1V3	09U1V3 caenorhabd1
45	32	76.2	479	2	086742	086742 streptomyce

ALIGNMENTS

RESULT 1
091163 PRELIMINARY; PRT; 66 AA.
ID 091163
AC 091163; 091164; 01, Created
DT 01-NOV-1996 (TRENBLER, 01, Last sequence update)
DT 01-JAN-1999 (TRENBLER, 09, Last sequence update)
DT 01-OCT-2000 (TRENBLER, 15, Last annotation update)
DE PREPROINSULIN (INSULIN B-CHAIN) (FRAGMENT).
OS Oncochinchus keta (Chum salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX NCBI_TaxID=8018;
RN [1]
RP SEQUENCE FROM N.A.
RA Kavan V.M.;
RL Submitted (SEP-1987) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR EMBL: K03487; AAA49415.1; -.
DR HSSP: P01315; 6INS.
DR InterPro: IPR000739; -.
DR Pfam: PF00049; Insulin; 1.
DR PRINTS: PR00276; INSULIN.
DR PROSITE: PS00262; INSULIN; 1.
FT NON_TER
SQ SEQUENCE 66 AA; 7676 MW; 8015B662D8BAD8DB CRC64;

Query Match /90.5%; Score 38; DB 13; Length 66;
Best local Similarity 85.7%; Pred. No. 1.6;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGERGFY 8
Db 1 GGERGFY 7
RESULT 2
ID 091807 PRELIMINARY; PRT; 106 AA.
AC 091807;

DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
 DE PREPROINSULIN.
 OS Rana pipiens (Northern leopard frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Neobatrachia; Ranioidea; -Ranidae; Rana.
 OX NCBI_TaxID=8404;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Irwin D.M., Stvarajah P.;
 RT "Proinsulin cDNAs from the leopard frog, Rana pipiens: evolution of
 RT proinsulin processing.";
 RL Comp. Biochem. Physiol. B, Biochem. Mol. Biol. 125:405-410(2000).
 CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
 DR EMBL: AF227187; AAF87285.1; -
 DR InterPro: IPR000739; -
 DR Pfam: PF00049; Insulin; 1.
 DR PRINTS: PR00276; INSULIN.
 DR PRINTS: PR00277; INSULIN.
 DR PROSITE: PS00262; INSULIN; 1.
 DR SMART: SM00078; IIGF; 1.
 SO SEQUENCE 106 AA; 12183 MW; 3A870EEC70217F92 CRC64;

Query Match 90.5%; Score 38; DB 13; Length 106;
 Best Local Similarity 85.7%; Pred. No. 2.6;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFEY 8
 DB 43 GDRGFEY 49

RESULT 3
 ID 022326 PRELIMINARY; PRT; 1259 AA.
 AC 022326;
 DT 01-JAN-1998 (TREMblrel. 05, Created)
 DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
 DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
 DE KINESIN-LIKE CALMODULIN-BINDING PROTEIN.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA Reddy A.S.N.;
 RL Gene 0:0-0(1997).
 DR EMBL: AF002220; AAC49901.1; -
 DR HSSP: P17119; 3KAR.
 DR Mendel: 24314; Arab.1078;24314.
 DR InterPro: IPR000299; -
 DR InterPro: IPR000515; -
 DR InterPro: IPR000857; -
 DR InterPro: IPR001752; -
 DR Pfam: PF00225; kinesin; 1.
 DR Pfam: PF00784; MYTH4; 1.
 DR PRINTS: PS00380; KINESINHEAVY.
 DR PROSITE: PS00402; BPD_TRANSF_INN_MEMBER; UNKNOWN_1.
 DR PROSITE: PS00411; KINESIN_MOTOR_DOMAIN1; UNKNOWN_1.
 DR PROSITE: PS0067; KINESIN_MOTOR_DOMAIN2; 1.
 DR SMART: SM00139; MYTH4; 1.
 KW ATP-binding; Coiled coil; Microtubules; Motor protein.
 SO SEQUENCE 1259 AA; 143320 MW; EAAAJA0A062C8838 CRC64;

Query Match 88.1%; Score 37; DB 10; Length 1259;

Best Local Similarity 85.7%; Pred. No. 65;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFEY 8
 DB 88 GKRGEFY 94

RESULT 4
 ID 023102 PRELIMINARY; PRT; 1259 AA.
 AC 023102;
 DT 01-JAN-1998 (TREMblrel. 05, Created)
 DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
 DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
 DE KINESIN-LIKE PROTEIN.
 GN ZWICHEL.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RX MEDLINE=97322360; PubMed=9177205;
 RA Oppenheimer D.G., Pollock M.A., Vacik J., Szymanski D.B., Ericson B.,
 RA Feldmann K., Marks M.D.;
 RT "Essential role of a kinesin-like protein in Arabidopsis trichome
 RT morphogenesis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 94:6261-6266(1997).
 DR EMBL: AF002678; AAB61712.1; -
 DR HSSP: P17119; 3KAR.
 DR Mendel: 24419; Arab.1078;24419.
 DR InterPro: IPR000299; -
 DR InterPro: IPR000515; -
 DR InterPro: IPR000857; -
 DR InterPro: IPR001752; -
 DR Pfam: PF00225; kinesin; 1.
 DR Pfam: PF00784; MYTH4; 1.
 DR PRINTS: PS00380; KINESINHEAVY.
 DR PROSITE: PS00402; BPD_TRANSF_INN_MEMBER; UNKNOWN_1.
 DR PROSITE: PS00411; KINESIN_MOTOR_DOMAIN1; UNKNOWN_1.
 DR PROSITE: PS0067; KINESIN_MOTOR_DOMAIN2; 1.
 DR SMART: SM00139; MYTH4; 1.
 KW ATP-binding; Coiled coil; Microtubules; Motor protein.
 SO SEQUENCE 1259 AA; 143318 MW; 95B41DF2B21BF26E CRC64;

Query Match 88.1%; Score 37; DB 10; Length 1259;
 Best Local Similarity 85.7%; Pred. No. 65;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFEY 8
 DB 88 GKRGEFY 94

RESULT 5
 ID 09FHN8 PRELIMINARY; PRT; 1260 AA.
 AC 09FHN8;
 DT 01-MAR-2001 (TREMblrel. 16, Created)
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
 DE KINESIN-LIKE CALMODULIN-BINDING PROTEIN.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-COLUMBIA;
 RX MEDLINE-20181125; PubMed-10718197;
 RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
 RT features of the regions of 3,076,755 bp covered by sixty P1 and TAC
 RT clones.";
 RL DNA Res. 7:31-63(2000).
 DR EMBL: AB018108; BAB11140.1; -;
 SQ SEQUENCE 1260 AA; 143448 MW; EE6B562CA6201EF CRC64;

Query Match 88.1%; Score 37; DB 10; Length 1260;
 Best Local Similarity 85.7%; Pred. No. 66;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GRCFFY 8
 DB 88 GRCFFY 94

RESULT 6
 ID 039130 PRELIMINARY; PRT; 1261 AA.
 AC 039130;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
 DE CALMODULIN-BINDING PROTEIN.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-96215138; PubMed-8636137;
 RA Reddy A.S.N., Safadi F., Narasimulu S.B., Golovkin M., Hu X.;
 RT "A novel plant calmodulin-binding protein with a kinesin heavy chain
 RT motor domain.";
 RL J. Biol. Chem. 271:7052-7060(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Reddy A.S.N.;
 RL Submitted (FEB-1996) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Reddy A.S.N.;
 RL Submitted (MAR-1996) to the EMBL/Genbank/DBJ databases.
 DR EMBL: L40358; AAC37475.1; -;
 DR HSP: P17119; 3KAR.
 DR Mendel: 6340; Arath.1078;6340.
 DR InterPro: IPR000239; -;
 DR InterPro: IPR000515; -;
 DR InterPro: IPR000857; -;
 DR InterPro: IPR001752; -;
 DR Pfam: PF00225; Kinesin; 1.
 DR Pfam: PF00784; MYTH4; 1.
 DR PRINTS: PR00380; KINESINHEAVY.
 DR PROSITE: PS50057; BAND_41_3; 1.
 DR PROSITE: PS00402; BPD_TRANSF_INN_MEMBR; UNKNOWN_1.
 DR PROSITE: PS00411; KINESIN_MOTOR_DOMAIN1; UNKNOWN_1.
 DR PROSITE: PS50067; KINESIN_MOTOR_DOMAIN2; 1.
 DR SMART: SM00139; MYTH4; 1.
 KW ATP-binding; Calmodulin-binding; Coiled coil; Microtubules;
 KW Motor protein.
 SQ SEQUENCE 1261 AA; 143367 MW; E9B52917652BCDCC CRC64;

Query Match 88.1%; Score 37; DB 10; Length 1261;
 Best Local Similarity 85.7%; Pred. No. 66;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GRCFFY 8
 DB 88 GRCFFY 94

RESULT 7
 ID 046795 PRELIMINARY; PRT; 141 AA.
 AC 046795;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-FEB-1997 (TREMblrel. 02, Last annotation update)
 DE ORF_F141.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K-12;
 RA Plunkett G.;
 RL Submitted (JUN-1995) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K-12;
 RA Blattner F.R., Plunkett G. III, Mayhew G.F., Perna N.T., Glasner F.D.;
 RL Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.
 DR EMBL: U28375; AAA83041.1; -;
 DR EMBL: AE000369; AAC75898.1; -;
 SQ SEQUENCE 141 AA; 16403 MW; CAEB421DA23E8681 CRC64;

Query Match 83.3%; Score 35; DB 2; Length 141;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 GRCFFY 8
 DB 34 GRCFFY 39

RESULT 8
 ID 09RDY9 PRELIMINARY; PRT; 188 AA.
 AC 09RDY9;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
 DE RMLC PROTEIN.
 GN RMLC.
 OS Legionella pneumophila.
 OC Bacteria; Proteobacteria; gamma subdivision; Legionellaceae group;
 OC Legionellaceae; Legionella.
 OX NCBI_TaxID=446;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RCL;
 RA Lueneberg E., Zetmann N., Hartmann M., Knirel Y.A., Kooststra O.,
 RA Zaehner U., Helbig J., Frosch M.;
 RT "A 30 kb gene cluster involved in biosynthesis of the virulence
 RT associated lipopolysaccharide carbohydrate moiety of Legionella
 RT pneumophila.";
 RL Submitted (JUN-1998) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AJ007311; CAB65208.1; -;
 DR InterPro: IPR000888; -;
 DR Pfam: PF00908; dTDP-sugar_isom; 1.
 SQ SEQUENCE 188 AA; 21509 MW; 42DA74D3D410E462 CRC64;

Query Match 83.3%; Score 35; DB 2; Length 188;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 ERGFFY 8
 DB 22 ERGFFY 27

RESULT 9
 ID 09PCX4 PRELIMINARY: PRT: 349 AA.
 AC 09PCX4;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE HYPOTHETICAL PROTEIN XF1630.
 GN XF1630.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 OC Xylella.
 OX NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=945C; PubMed=10910347;
 RX MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
 Alvaranga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,
 Barros M.H., Bonaccorsi E.D., Bordin S., Boye J.M., Britones M.R.S.,
 Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,
 Colauto N.B., Colombo C., Costa F.P., Costa M.C.R., Costa-Neto C.M.,
 Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
 Fagundes A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,
 Garner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 Ho P.L., Hohlseil J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 Klieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 Machado M.A., Madeira A.M.B.N., Martins E.M.F., Matsukuma A.Y.,
 Marques M.V., Marzaca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 Menck C.F.M., Miracca E.C., Nunes L.R., Oliveira M.A.,
 Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 da Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 Peixoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pesquero J.B.,
 Quagga R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 de Rosa V.E.Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A.Jr.,
 da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
 Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 Zago M.A., Zatz M., Meidanis J., Secubal J.C.;
 RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
 RL Nature 406:151-159(2000).
 DR EMBL, AE003990, AAF84439.1;
 KW Hypothetical protein.
 SQ SEQUENCE 349 AA; 39805 MW; B5E1FBCE740180BE CRC64;

Query Match 83.3%; Score 35; DB 2; Length 349;
 Best Local Similarity 85.7%; Pred. NO. 39;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ERGFFY 8
 DB 68 ERGFFY 74

RESULT 10
 ID 043607 PRELIMINARY: PRT: 551 AA.
 AC 043607;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE PRUNIN PRECURSOR.

GN PRUN.
 OS Prunus dulcis (Almond) (Prunus amygdalus).
 OC Eukaryota; Viridiplantae; Embryophyta; tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eucotsids I;
 OC Rosales; Rosaceae; Prunus.
 OX NCBI_TaxID=3755;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV, TEXAS; TISSUE=IMMATURE SEEDLING;
 RX MEDLINE=95170003; PubMed=7865791;
 RA Garcia-Mas J., Messegue R., Arus P., Puigdomenech P.;
 RT "Molecular characterization of cDNAs corresponding to genes expressed
 during almond (Prunus amygdalus Batsch) seed development.";
 RL Plant Mol. Biol. 27:205-210(1995).
 CC -1- SUBUNIT: HEXAMER, EACH SUBUNIT IS COMPOSED OF AN ACIDIC AND A
 BASIC CHAIN DERIVED FROM A SINGLE PRECURSOR AND LINKED BY A
 DISULFIDE BOND (BY SIMILARITY).
 CC -1- SIMILARITY: TO OTHER 11S SEED STORAGE PROTEINS (GLOBULINS).
 CC EMBL: X78119; CAA55009.1; -.
 DR InterPro: IPR000459; -.
 DR Pfam: PF00190; Seedstore_11s; 2.
 DR PRINTS: PR00439; 11SGLOBULIN
 DR PROSITE: PS00305; 11S_SEED_STORAGE; 1.
 KW Multigene family; Signal.
 FT SIGNAL 1
 SQ SEQUENCE 551 AA; 63017 MW; 70D93418A222BA8F CRC64;

Query Match 83.3%; Score 35; DB 10; Length 551;
 Best Local Similarity 100.0%; Pred. NO. 66;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 ERGFFY 8
 DB 417 ERGFFY 422

RESULT 11
 ID P91871 PRELIMINARY: PRT: 2586 AA.
 AC P91871;
 DT 01-MAY-1997 (TREMBLrel. 03, Created)
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE F32H2.5 PROTEIN.
 GN F32H2.5.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kershaw J.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
 Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 Watson A., Weinstock L., Wilkinson-Sproat J., Woldman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 RL Nature 368:32-38(1994).
 DR EMBL: Z81523; CAB04244.1; -.
 DR InterPro: IPR000255; -.
 DR InterPro: IPR000794; -.

DR InterPro: IPR001031; -
 DR InterPro: IPR001227; -
 DR InterPro: IPR002085; -
 DR InterPro: IPR003006; -
 DR Pfam: PF00107; adh_zinc; 1.
 DR Pfam: PF00109; ketoacyl-synt; 1.
 DR Pfam: PF00550; pp-binding; 1.
 DR Pfam: PF00698; Acyl_transf; 1.
 DR Pfam: PF00975; Thioesterase; 1.
 DR PROSITE: PS00075; ACP_DOMAIN; 1.
 DR PROSITE: PS00606; B_KETOACYL-SYNTASE; 1.
 DR PROSITE: PS00290; IG_MHC; UNKNOWN; 1.
 DR Phosphatetheine; transferase.
 SO SEQUENCE 2586 AA; 286378 MW; E97D658836FCA674 CRC64;

Query Match 83.3%; Score 35; DB 5; Length 2586;
 Best Local Similarity 85.7%; Pred. No. 3; e+02;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GERGFFX 8
 DB 1035 GERGFFX 1041

RESULT 12
 O91161 PRELIMINARY; PRT; 116 AA.
 AC O91161;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE INSULIN-LIKE GROWTH FACTOR I PRECURSOR (FRAGMENT).
 OS Oncorhynchus kisutch (Coho salmon).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
 OX NCBI_TaxID=8019;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LIVER;
 RX MEDLINE=90190659; PubMed=2628735;
 RA Cao Q.P., Duguay S.J., Pilsetskaya E., Steiner D.F., Chan S.J.;
 RT "Nucleotide sequence and growth hormone-regulated expression of salmon
 insulin-like growth factor I mRNA.";
 RL Mol. Endocrinol. 3:2005-2010(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LIVER;
 RX MEDLINE=93024477; PubMed=1406698;
 RA Duguay S.J., Park L.K., Samadpour M., Dickhoff W.W.;
 RT "Nucleotide sequence and tissue distribution of three insulin-like
 growth factor I prohormones in salmon.";
 RL Mol. Endocrinol. 6:1202-1210(1992).
 CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
 DR EMBL: M81911; AAB59947.1; -
 DR HSSP: P05019; IGFL
 DR InterPro: IPR000739; -
 DR Pfam: PF00049; Insulin; 1.
 DR PRINTS: PR00276; INSULIN.
 DR PRINTS: PR00277; INSULIN.
 DR PROSITE: PS00262; INSULIN; 1.
 DR SMART: SM00078; IIGF; 1.
 KM Signal.
 FT NON_TER 1 1
 FT SIGNAL <1 18 POTENTIAL.
 FT CHAIN 19 88 INSULIN-LIKE GROWTH FACTOR I.
 FT NON_TER 116 116
 SO SEQUENCE 116 AA; 12697 MW; C5F378915179D89D CRC64;

Query Match 81.0%; Score 34; DB 13; Length 116;

Best Local Similarity 71.4%; Pred. No. 18;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFFX 8
 DB 37 GERGFFX 43

RESULT 13
 O91476 PRELIMINARY; PRT; 117 AA.
 AC O91476;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE INSULIN-LIKE GROWTH FACTOR I PRECURSOR (FRAGMENT).
 OS Salmo salar (Atlantic salmon).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
 OX NCBI_TaxID=8030;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LIVER;
 RX MEDLINE=93024477; PubMed=1406698;
 RA Duguay S.J., Park L.K., Samadpour M., Dickhoff W.W.;
 RT "Nucleotide sequence and tissue distribution of three insulin-like
 growth factor I prohormones in salmon.";
 RL Mol. Endocrinol. 6:1202-1210(1992).
 CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
 DR EMBL: M81904; AAA18212.1; -
 DR HSSP: P05019; IGFL
 DR InterPro: IPR000739; -
 DR Pfam: PF00049; Insulin; 1.
 DR PRINTS: PR00276; INSULIN.
 DR PRINTS: PR00277; INSULIN.
 DR PROSITE: PS00262; INSULIN; 1.
 DR SMART: SM00078; IIGF; 1.
 KM Signal.
 FT NON_TER 1 1
 FT SIGNAL <1 18 POTENTIAL.
 FT CHAIN 19 88 INSULIN-LIKE GROWTH FACTOR I.
 SO SEQUENCE 117 AA; 12867 MW; A97666EE2F526EAC CRC64;

Query Match 81.0%; Score 34; DB 13; Length 117;
 Best Local Similarity 71.4%; Pred. No. 19;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFFX 8
 DB 37 GERGFFX 43

RESULT 14
 O91475 PRELIMINARY; PRT; 145 AA.
 AC O91475;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE INSULIN-LIKE GROWTH FACTOR I PRECURSOR (FRAGMENT).
 OS Salmo salar (Atlantic salmon).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
 OX NCBI_TaxID=8030;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LIVER;
 RX MEDLINE=93024477; PubMed=1406698;
 RA Duguay S.J., Park L.K., Samadpour M., Dickhoff W.W.;

RT *Nucleotide sequence and tissue distribution of three insulin-like
 RT growth factor I prohormones in salmon.";
 RL Mol. Endocrinol. 6:1202-1210(1992).
 CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
 DR EMBL: M81904; AAA18211.1; -.
 DR HSSP: P05019; IGFI.
 DR InterPro: IPR000739; -.
 DR Pfam: PF00049; Insulin; 1.
 DR PRINTS: PR00276; INSULINA.
 DR PRINTS: PR00277; INSULINB.
 DR PROSITE: PS00262; INSULIN; 1.
 DR SMART: SM00078; IIGF; 1.
 DR Signal.
 KW NON_TER 1 1
 FT SIGNAL <1 18 POTENTIAL.
 FT CHAIN 19 88 INSULIN-LIKE GROWTH FACTOR I.
 FT NON_TER 145 145
 SO SEQUENCE 145 AA; 15885 MW; 3D94EDF477268FC4 CRC64;

Query Match 81.0%; Score 34; DB 13; Length 145;
 Best Local Similarity 71.4%; Pred. No. 24;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFY 8
 |||||:
 DB 37 GERGFY 43

RESULT 15
 ID 091231 PRELIMINARY; PRT; 149 AA.
 AC 091231;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
 DE INSULIN-LIKE GROWTH FACTOR-I.
 GN IGF-I.
 OS Oncorhynchus tshawytscha (Chinook salmon) (King salmon).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 OC Proacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
 OX NCBI_TaxID=74940;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BIG QUALICUM RIVER, B.C.; TISSUE=LIVER;
 RX MEDLINE=93247592; PubMed=7683374;
 RA Wallis A.E.; Devlin R.H.;
 RT "Duplicate insulin-like growth factor-I genes in salmon display
 RT alternative splicing pathways";
 RL Mol. Endocrinol. 7:409-422(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BIG QUALICUM RIVER, B.C.; TISSUE=LIVER;
 RA Devlin R.H.;
 RL Submitted (Oct-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
 CC EMBL: U15962; AAA67268.1; -.
 DR HSSP: P05019; IGFI.
 DR InterPro: IPR000739; -.
 DR Pfam: PF00049; Insulin; 1.
 DR PRINTS: PR00276; INSULINA.
 DR PRINTS: PR00277; INSULINB.
 DR ProDom: PD001048; -; 1.
 DR PROSITE: PS00262; INSULIN; 1.
 DR SMART: SM00078; IIGF; 1.
 SO SEQUENCE 149 AA; 16507 MW; 9AC8F072762D2AA0 CRC64;

Query Match 81.0%; Score 34; DB 13; Length 149;
 Best Local Similarity 71.4%; Pred. No. 24;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFY 8
 |||||:
 DB 63 GERGFY 69

Search completed: September 6, 2001, 16:49:52
 Job time: 736 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:45:51 : Search time 134.15 Seconds
(without alignments)
4.543 Million cell updates/sec

Title: US-09-603-713-23

Perfect score: 42

Sequence: 1 XGERGFFY 8

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR_68:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	97.6	17	2	INS5778
2	41	97.6	50	1	INBN2
3	41	97.6	51	1	INWHP
4	41	97.6	51	1	INWHP
5	41	97.6	51	1	INWHP
6	41	97.6	51	1	INWHP
7	41	97.6	51	1	INWHP
8	41	97.6	51	1	INWHP
9	41	97.6	51	1	INWHP
10	41	97.6	51	1	INWHP
11	41	97.6	51	1	INWHP
12	41	97.6	51	1	INWHP
13	41	97.6	51	1	INWHP
14	41	97.6	51	1	INWHP
15	41	97.6	51	1	INWHP
16	41	97.6	51	1	INWHP
17	41	97.6	51	1	INWHP
18	41	97.6	51	1	INWHP
19	41	97.6	51	1	INWHP
20	41	97.6	51	1	INWHP
21	41	97.6	51	1	INWHP
22	41	97.6	51	1	INWHP
23	41	97.6	51	1	INWHP
24	41	97.6	51	1	INWHP
25	41	97.6	51	1	INWHP
26	41	97.6	51	1	INWHP
27	41	97.6	51	1	INWHP
28	41	97.6	51	1	INWHP
29	41	97.6	51	1	INWHP

30	41	97.6	86	1	IPHO
31	41	97.6	96	2	PC7082
32	41	97.6	103	2	IS1221
33	41	97.6	105	1	IPBO
34	41	97.6	107	1	IPCH
35	41	97.6	108	1	INMS1
36	41	97.6	108	2	S09278
37	41	97.6	108	2	A39883
38	41	97.6	110	1	IPHU
39	41	97.6	110	1	INRB
40	41	97.6	110	1	IPDG
41	41	97.6	110	1	IPRT1
42	41	97.6	110	1	IPRT2
43	41	97.6	110	1	INMS2
44	41	97.6	110	2	B42179
45	41	97.6	110	2	A42179

ALIGNMENTS

RESULT 1

INS5778

Insulin chain B - bovine (fragments)

C:Species: Bos primigenius indicus x Bos primigenius taurus (cattle)

C>Date: 19-Mar-1997 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999

C:Accession: S15778; S15779

R:Bertram, T.; Agerberth, B.; Joernvall, H.

FEBS Lett. 283, 100-103, 1991

A>Title: Direct analysis of peptides and amino acids from capillary electrophoresis.

A:Reference number: S15778; MUID:91243852

A:Accession: S15778

A:Status: preliminary

A:Molecule type: protein

A:Residues: 1-8 <PEB1>

A:Accession: S15779

A:Status: preliminary

A:Molecule type: protein

A:Residues: 9-17 <PEB2>

C:Superfamily: Insulin

C:Keywords: hormone; pancreas

Query Match

Best Local Similarity 97.6%; Score 41; DB 2; Length 17;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GERGFFY 8

DB 7 GERGFFY 13

RESULT 2

INBN2

Insulin - skipjack tuna (tentative sequence)

C:Species: Euthynnus pelamis, Katsuwonus pelamis (skipjack tuna)

C>Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change 31-Mar-2000

C:Accession: A01607; A91916

R:Kotaki, A.

J. Biochem. 53, 61-70, 1963

A>Title: Studies on Insulin. V. On the structure of the glycy chain of bonito Insulin

A:Reference number: A91917

A:Accession: A01607

A:Molecule type: protein

A:Residues: 1-29 <KOT1>

R:Kotaki, A.

J. Biochem. 51, 301-309, 1962

A>Title: Studies on Insulin. III. On the structure of the alanyl chain of bonito Insu

A:Reference number: A91916

A:Accession: A91916

A:Molecule type: protein

A:Residues: 30-50 <KOT2>

C:Superfamily: Insulin

C:Keywords: hormone; pancreas
F:1-29/Domain: Insulin chain B #status experimental <BCH>
F:1-29,30-50/Product: Insulin #status experimental <MAT>
F:30-50/Domain: Insulin chain A #status experimental <ACH>
F:7-36,19-49,35-40/Disulfide bonds: #status experimental

Query Match 97.6%; Score 41; DB 1; Length 50;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFY 8
|||||||
Db 20 GERGFY 26

RESULT 3

INMHP
Insulin - sperm whale
C:Species: Physeter catodon (sperm whale)
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 16-Jul-1999
C:Accession: A93142; A90082
R: Ishihara, Y.; Saito, T.; Ito, Y.; Fujino, M.
Nature 181, 1468-1469, 1958
A:Title: Structure of sperm- and sei-whale insulins and their breakdown by whale pepsin.
A:Reference number: A93142
A:Accession: A93142
A:Molecule type: protein
A:Residues: 1-30,31-51 <ISH>
R: Harris, J.L.; Sanger, F.; Naughton, M.A.
Arch. Biochem. Biophys. 65, 427-428, 1956
A:Title: Species differences in insulin.
A:Reference number: A90082
A:Accession: A90082
A:Molecule type: protein
A:Residues: 1-30,31-51 <HAR>
C:Superfamily: Insulin
C:Keywords: hormone; pancreas
F:1-30/Domain: Insulin chain B #status experimental <BCH>
F:1-30,31-51/Product: Insulin #status experimental <MAT>
F:31-51/Domain: Insulin chain A #status experimental <ACH>
F:7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 97.6%; Score 41; DB 1; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFY 8
|||||||
Db 20 GERGFY 26

RESULT 4

INMHP
Insulin - finback whale (tentative sequence)
C:Species: Balenoptera physalus (finback whale, common rorqual)
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 16-Jul-1999
C:Accession: A91918
R: Hama, H.; Titani, K.; Sakaki, S.; Narita, K.
J. Biochem. 56, 285-293, 1964
A:Title: The amino acid sequence in fin-whale insulin.
A:Reference number: A91918
A:Accession: A91918
A:Molecule type: protein
A:Residues: 1-30,31-51 <HAM>
C:Superfamily: Insulin
C:Keywords: hormone; pancreas
F:1-30/Domain: Insulin chain B #status experimental <BCH>
F:1-30,31-51/Product: Insulin #status experimental <MAT>
F:31-51/Domain: Insulin chain A #status experimental <ACH>
F:7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 97.6%; Score 41; DB 1; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFY 8
|||||||
Db 20 GERGFY 26

RESULT 5

INMHS
Insulin - sei whale
C:Species: Balenoptera borealis (sei whale)
C:Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change 16-Jul-1999
C:Accession: A01582
R: Ishihara, Y.; Saito, T.; Ito, Y.; Fujino, M.
Nature 181, 1468-1469, 1958
A:Title: Structure of sperm- and sei-whale insulins and their breakdown by whale pepsin.
A:Reference number: A93142
A:Accession: A01582
A:Molecule type: protein
A:Residues: 1-30,31-51 <ISH>
C:Superfamily: Insulin
C:Keywords: hormone; pancreas
F:1-30/Domain: Insulin chain B #status experimental <BCH>
F:1-30,31-51/Product: Insulin #status experimental <MAT>
F:31-51/Domain: Insulin chain A #status experimental <ACH>
F:7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 97.6%; Score 41; DB 1; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFY 8
|||||||
Db 20 GERGFY 26

RESULT 6

INEL
Insulin - elephant
C:Species: Elephantidae gen. sp. (elephant)
C:Date: 24-Apr-1984 #sequence_revision 30-Sep-1988 #text_change 16-Jul-1999
C:Accession: A01584
R: Smith, L.F.
Am. J. Med. 40, 662-666, 1966
A:Title: Species variation in the amino acid sequence of insulin.
A:Reference number: A90029; MUID:66160119
A:Accession: A01584
A:Molecule type: protein
A:Residues: 1-30,31-51 <SMI>
A:Note: The species of elephant is not given, but it is most probably the Indian elep
C:Superfamily: Insulin
C:Keywords: hormone; pancreas
F:1-30/Domain: Insulin chain B #status experimental <BCH>
F:1-30,31-51/Product: Insulin #status experimental <MAT>
F:31-51/Domain: Insulin chain A #status experimental <ACH>
F:7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 97.6%; Score 41; DB 1; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFY 8
|||||||
Db 20 GERGFY 26

RESULT 7

INGT

Insulin - goat
C:Species: Capra aegagrus hircus (domestic goat)
C:Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change 16-Jul-1999
C:Accession: A01586
R:Smith, L.F.
Am. J. Med. 40, 662-666, 1966
A:Title: Species variation in the amino acid sequence of Insulin.
A:Reference number: A90029; MUID:66160119
A:Accession: A01586
A:Molecule type: protein
A:Residues: 1-30:31-51 <SMI>
C:Superfamily: Insulin
C:Keywords: hormone; pancreas
F:1-30/Domain: Insulin chain B #status experimental <BCH>
F:1-30,31-51/Product: Insulin #status experimental <MAT>
F:31-51/Domain: Insulin chain A #status experimental <ACH>
F:7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 97.6%; Score 41; DB 1; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFY 8
|||||
DB 20 GERGFY 26

RESULT 8
INCM
Insulin - Arabian camel (tentative sequence)
C:Species: Camelus dromedarius (Arabian camel)
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 16-Jul-1999
C:Accession: A92782
R:Danho, W.O.
J. Fac. Med. Baghdad 14, 16-28, 1972
A:Title: The isolation and characterization of Insulin of camel (Camelus dromedarius).
A:Reference number: A92782
A:Accession: A92782
A:Molecule type: protein
A:Residues: 1-30:31-51 <DAN>
C:Superfamily: Insulin
C:Keywords: hormone; pancreas
F:1-30/Domain: Insulin chain B #status experimental <BCH>
F:1-30,31-51/Product: Insulin #status experimental <MAT>
F:31-51/Domain: Insulin chain A #status experimental <ACH>
F:7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 97.6%; Score 41; DB 1; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFY 8
|||||
DB 20 GERGFY 26

RESULT 9
INCT
Insulin - cat
C:Species: Felis silvestris catus (domestic cat)
C:Date: 30-Jun-1987 #sequence_revision 30-Sep-1988 #text_change 16-Jul-1999
C:Accession: A01588
R:Halliden, G.; Gafvelin, G.; Mutt, V.; Jornvall, H.
Arch. Biochem. Biophys. 247, 20-27, 1986
A:Title: Characterization of cat Insulin.
A:Reference number: A01588; MUID:86214076
A:Accession: A01588
A:Molecule type: protein
A:Residues: 1-30:31-51 <HAL>
C:Superfamily: Insulin
C:Keywords: hormone; pancreas

F:1-30/Domain: Insulin chain B #status experimental <BCH>
F:1-30,31-51/Product: Insulin #status experimental <MAT>
F:31-51/Domain: Insulin chain A #status experimental <ACH>
F:7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 97.6%; Score 41; DB 1; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFY 8
|||||
DB 20 GERGFY 26

RESULT 10
INHY
Insulin - hamster
C:Species: Cricetinae gen. sp. (hamster)
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 16-Jul-1999
C:Accession: A91456
R:Neelon, F.A.; Delcher, H.K.; Steilman, H.; Lebovitz, H.E.
Fed. Proc. 32, 300, 1973
A:Title: Structure of hamster Insulin: comparison with a tumor Insulin.
A:Reference number: A91456
A:Accession: A91456
A:Molecule type: protein
A:Residues: 1-30:31-51 <NEE>
C:Superfamily: Insulin
C:Keywords: hormone; pancreas
F:1-30/Domain: Insulin chain B #status experimental <BCH>
F:1-30,31-51/Product: Insulin #status experimental <MAT>
F:31-51/Domain: Insulin chain A #status experimental <ACH>
F:7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 97.6%; Score 41; DB 1; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFY 8
|||||
DB 20 GERGFY 26

RESULT 11
INMSSP
Insulin - Egyptian spiny mouse (tentative sequence)
C:Species: Acomys calhirus (Egyptian spiny mouse)
C:Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change 31-Mar-2000
C:Accession: A01591
R:Buenzli, H.F.; Humbel, R.E.
Hoppe-Seyler's Z. Physiol. Chem. 353, 444-450, 1972
A:Title: Isolation and partial structural analysis of Insulin from mouse (Mus musculus)
A:Reference number: A01591; MUID:72189454
A:Contents: composition
A:Accession: A01591
A:Molecule type: protein
A:Residues: 1-30:31-51 <BUE>
C:Superfamily: Insulin
C:Keywords: hormone; pancreas
F:1-30/Domain: Insulin chain B #status predicted <BCH>
F:1-30,31-51/Product: Insulin #status predicted <MAT>
F:31-51/Domain: Insulin chain A #status predicted <ACH>
F:7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 97.6%; Score 41; DB 1; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFY 8
|||||

Db 20 GERGFFY 26

RESULT 12

INMKSO
Insulin - common squirrel monkey
C/Species: Saimiri sciureus (common squirrel monkey)
C/Date: 18-Oct-1991 #sequence_revision 02-May-1994 #text_change 16-Jul-1999
C/Accession: B39258; A39258
R.Yu, J.H.; Eng, J.; Yalow, R.S.
Proc. Natl. Acad. Sci. U.S.A. 87, 9766-9768, 1990
A/Title: Isolation and amino acid sequences of squirrel monkey (Saimiri sciurea) insulin
A/Reference number: A39258; MUID:91088593
A/Accession: B39258
A/Molecule type: protein
A/Residues: 1-30 <YUB>
A/Accession: A39258
A/Molecule type: protein
A/Residues: 31-51 <YUA>
C/Superfamily: insulin
C/Keywords: hormone; pancreas
F:1-30/Domain: insulin chain B #status experimental <BCH>
F:1-30/31-51/Product: insulin #status experimental <MAT>
F:31-51/Domain: insulin chain A #status experimental <ACH>
F:7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match

97.6%; Score 41; DB 1; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.12;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFFY 8

Db 20 GERGFFY 26

RESULT 13

INTK
Insulin -- turkey (tentative sequence)
C/Species: Meleagris gallopavo (common turkey)
C/Date: 22-Jun-1981 #sequence_revision 22-Jun-1981 #text_change 31-Mar-2000
C/Accession: A01599
R.Jentsch, J.
Submitted to the Atlas, February 1970
A/Reference number: A01599
A/Accession: A01599
A/Molecule type: protein
A/Residues: 1-30;31-51 <JEN>
C/Superfamily: insulin
C/Keywords: hormone; pancreas
F:1-30/Domain: insulin chain B #status experimental <BCH>
F:1-30,31-51/Product: insulin #status experimental <MAT>
F:31-51/Domain: insulin chain A #status experimental <ACH>
F:7-37,19-50,36-41/Disulfide bonds: #status experimental

Query Match

97.6%; Score 41; DB 1; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.12;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFFY 8

Db 20 GERGFFY 26

RESULT 14

INOS
Insulin - ostrich
C/Species: Struthio camelus (ostrich)
C/Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 16-Jul-1999
C/Accession: JK0017
B.Evans, T.K.; Lithauer, D.; Oelofsen, W.
Int. J. Pept. Protein Res. 31, 454-462, 1988

A/Title: Purification and primary structure of ostrich insulin.

A/Reference number: JK0017; MUID:88514456

A/Accession: JK0017

A/Molecule type: protein

A/Residues: 1-30;31-51 <INS>

C/Superfamily: insulin

C/Keywords: hormone; pancreas

F:1-30/Domain: insulin chain B #status experimental <BCH>

F:1-30,31-51/Product: insulin #status experimental <MAT>

F:31-51/Domain: insulin chain A #status experimental <ACH>

F:7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match

97.6%; Score 41; DB 1; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.12;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFFY 8

Db 20 GERGFFY 26

RESULT 15

A61129
Insulin - black-bellied slider
C/Species: Chrysemys dorsibini (black-bellied slider)
C/Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 16-Jul-1999
C/Accession: A61129
R.Cascone, O.; Turyn, D.; Dellacha, J.M.; Machado, V.L.A.; Marques, M.; Vite, N.; Cas
Gen. Comp. Endocrinol. 84, 355-359, 1991
A/Title: Isolation, purification, and primary structure of insulin from the turtle Ch
A/Reference number: A61129; MUID:9225302
A/Accession: A61129
A/Molecule type: protein
A/Residues: 1-30;31-51 <CAS>
C/Superfamily: insulin
C/Keywords: hormone; pancreas
F:1-30/Domain: insulin chain B #status experimental <BCH>
F:1-30,31-51/Product: insulin #status experimental <MAT>
F:31-51/Domain: insulin chain A #status experimental <ACH>
F:7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match

97.6%; Score 41; DB 1; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.12;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFFY 8

Db 20 GERGFFY 26

Search completed: September 6, 2001, 16:45:51
Job time: 495 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:51:08 ; Search time 72.75 Seconds
(without alignments)
3.767 Million cell updates/sec

Title: US-09-603-713-23

Perfect score: 42
Sequence: 1 XGERGFY 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	97.6	50	1	INS_KATPE
2	41	97.6	51	1	INS_ACOCA
3	41	97.6	51	1	INS_ALMI
4	41	97.6	51	1	INS_ANSAM
5	41	97.6	51	1	INS_BALBO
6	41	97.6	51	1	INS_BALPH
7	41	97.6	51	1	INS_CAMDR
8	41	97.6	51	1	INS_CAPHT
9	41	97.6	51	1	INS_DIDMA
10	41	97.6	51	1	INS_ELEMA
11	41	97.6	51	1	INS_FELCA
12	41	97.6	51	1	INS_PLAFE
13	41	97.6	51	1	INS_TRASC
14	41	97.6	52	1	INS_ACTIG
15	41	97.6	59	1	INS_HIDCO
16	41	97.6	81	1	INS_ANAPL
17	41	97.6	86	1	INS_HORSE
18	41	97.6	89	1	INS_CALMI
19	41	97.6	103	1	INS_SEFRE
20	41	97.6	105	1	INS_BOVIN
21	41	97.6	105	1	INS_SHEEP
22	41	97.6	107	1	INS_CHICK
23	41	97.6	108	1	INS_MOUSE
24	41	97.6	108	1	INS_AOTTR
25	41	97.6	108	1	INS_PIG
26	41	97.6	108	1	INS_RODSP
27	41	97.6	110	1	INS_RAT
28	41	97.6	110	1	INS_MOUSE
29	41	97.6	110	1	INS2_RAT
30	41	97.6	110	1	INS_CANFA
31	41	97.6	110	1	INS_CERAE
32	41	97.6	110	1	INS_CRITO
33	41	97.6	110	1	INS_HUMAN

34	41	97.6	110	1	INS_MACFA	P30406 macaca fasc
35	41	97.6	110	1	INS_PANTR	P30410 pan troglod
36	41	97.6	110	1	INS_PSAOB	O62587 psammomys o
37	41	97.6	110	1	INS_RABIT	P01311 oryctolagus
38	41	97.6	115	1	INS_VERMO	O94712 verasper mo
39	38	90.5	50	1	INS2_BATSP	P01338 batrachoidi
40	38	90.5	50	1	INS_MYOSC	P07453 myoxocephal
41	38	90.5	50	1	INS_ONCGO	P23187 oncorhynchu
42	38	90.5	51	1	INS1_BATSP	P01337 batrachoidi
43	38	90.5	51	1	INS2_THUTH	P01339 thunnus thy
44	38	90.5	51	1	INS_CHIBR	P01327 chinchilla
45	38	90.5	51	1	INS_ZAODH	P12708 zaocys dhum

ALIGNMENTS

RESULT	ID	INS_KATPE	STANDARD	PRT	50 AA
AC	P01340	INS_KATPE			
DT	21-JUL-1986 (Rel. 01, Created)				
DT	21-JUL-1986 (Rel. 01, Last sequence update)				
DT	01-OCT-2000 (Rel. 40, Last annotation update)				
DE	INSULIN.				
GN	INS.				
OS	Katsuwonus pelamis (Skipjack tuna) (Bontco).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;				
OC	Acanthomorpha; Acanthopterygii; Perciformes; Scombroidei;				
OX	Scombridae; Euthynnus.				
NCBI_TaxID=8226;					
RP	SEQUENCE OF 1-29.				
RA	Kotaki A.;				
RT	"Studies on Insulin. V. On the structure of the glycy chain of				
RT	bontco insulin II.";				
RL	J. Biochem. 53:61-70(1963).				
RN	[2]				
RP	SEQUENCE OF 30-50.				
RA	Kotaki A.;				
RT	"Studies on Insulin. III. On the structure of the alanyl chain of				
RT	bontco insulin.";				
RL	J. Biochem. 51:301-309(1962).				
CC	-1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT				
CC	INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND				
CC	FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE				
CC	CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.				
CC	-1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO				
CC	DISULFIDE BONDS.				
CC	-1- SUBCELLULAR LOCATION: SECRETED.				
CC	-1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.				
CC	PIR: A01607; INBN2.				
DR	HSPR: P01308; ILPH.				
DR	InterPro: IPR000739; -				
DR	Pfam: PF00049; Insulin; 1.				
DR	PRINTS: PR00277; INSULIN.				
DR	PROSITE: PS00262; INSULIN; 1.				
KW	Insulin family; Hormone; Glucose metabolism.				
FT	CHAIN 1				
FT	NON-CONS 29				
FT	CHAIN 30				
FT	DISULFID 7				
FT	DISULFID 19				
FT	DISULFID 35				
FT	DISULFID 40				
FT	SEQUENCE 50 AA: 5697 MW: 3627578FE24CE92E CRC64:				

Query Match 97.6%; Score 41; DB 1; Length 50;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGERGFY 8

```
DB      20 GERGFFY 26

RESULT 2
INS_ACOCA  STANDARD:      PRT:      51 AA.
ID  INS_ACOCA  P01324;
AC  21-JUL-1986 (Rel. 01, Created)
DT  21-JUL-1986 (Rel. 01, Last sequence update)
DE  01-NOV-1997 (Rel. 35, Last annotation update)
DE  INSULIN.
GN  INS.
OS  Acomys cahirinus (Egyptian spiny mouse).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Acomys.
OX  NCBI_TaxId=10068;
RN  [1]
RP  COMPOSITION.
RX  MEDLINE=72189454; PubMed=5028210;
RA  Buenzli H.F.; Humbel R.E.;
RT  "Isolation and partial structural analysis of insulin from mouse (Mus musculus) and spiny mouse (Acomys cahirinus).";
RL  Hope-Seyler's Z. Physiol. Chem. 353:444-450(1972).
CC  -I- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC  -I- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO DISULFIDE BONDS.
CC  -I- SUBCELLULAR LOCATION: SECRETED.
CC  -I- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR  PIR: A01591; INMSP.
DR  HSSP: P01308; IREN.
DR  InterPro: IPR000739;
DR  Pfam: PF00049; Insulin; 1.
DR  PRINTS: PR00276; INSULIN.
DR  PRINTS: PR00277; INSULIN.
DR  PROSITE: PS00262; INSULIN; 1.
KW  Insulin family; Hormone; Glucose metabolism.
FT  CHAIN 1 30 INSULIN B CHAIN.
FT  NON_CONS 30 31
FT  CHAIN 31 51
FT  DISULFID 7 37 INTERCHAIN (BY SIMILARITY).
FT  DISULFID 19 50 INTERCHAIN (BY SIMILARITY).
FT  DISULFID 36 41 BY SIMILARITY.
FT  SEQUENCE 51 AA; 5768 MW; 992BD8B629047DD3 CRC64;
SQ

Query Match      97.6%; Score 41; DB 1; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 GERGFFY 8
DB      20 GERGFFY 26

RESULT 3
INS_ALIMI  STANDARD:      PRT:      51 AA.
ID  INS_ALIMI  P12703;
AC  01-OCT-1989 (Rel. 12, Created)
DT  01-NOV-1991 (Rel. 20, Last sequence update)
DE  01-OCT-2000 (Rel. 40, Last annotation update)
DE  INSULIN.
GN  INS.
OS  Alligator mississippiensis (American alligator).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Archosauria; Crocodylidae; Alligatorinae; Alligator.
OX  NCBI_TaxId=8496;
RN  [1]
RP  SEQUENCE.
```

```
RX  MEDLINE=84262419; PubMed=6146554;
RA  Lance V.; Hamilton J.W.; Rouse J.B.; Kimmel J.R.; Pollock H.G.;
RT  "Isolation and characterization of reptilian insulin, glucagon, and pancreatic polypeptide: complete amino acid sequence of alligator (Alligator mississippiensis) insulin and pancreatic polypeptide.";
RL  Gen. Comp. Endocrinol. 55:112-124(1984).
CC  -I- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC  -I- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO DISULFIDE BONDS.
CC  -I- SUBCELLULAR LOCATION: SECRETED.
CC  -I- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR  PIR: S07210; INAO.
DR  HSSP: P01308; IREN.
DR  InterPro: IPR000739;
DR  Pfam: PF00049; Insulin; 1.
DR  PRINTS: PR00276; INSULIN.
DR  PRINTS: PR00277; INSULIN.
DR  PROSITE: PS00262; INSULIN; 1.
KW  Insulin family; Hormone; Glucose metabolism.
FT  CHAIN 1 30 INSULIN B CHAIN.
FT  NON_CONS 30 31
FT  CHAIN 31 51 INSULIN A CHAIN.
FT  DISULFID 7 37 INTERCHAIN.
FT  DISULFID 19 50 INTERCHAIN.
FT  DISULFID 36 41
FT  SEQUENCE 51 AA; 5689 MW; 19079A8B9879876A CRC64;
SQ

Query Match      97.6%; Score 41; DB 1; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 GERGFFY 8
DB      20 GERGFFY 26

RESULT 4
INS_ANSAN  STANDARD:      PRT:      51 AA.
ID  INS_ANSAN  P07454; 010995;
AC  P07454; 010995;
DT  01-APR-1988 (Rel. 07, Created)
DE  01-APR-1988 (Rel. 07, Last sequence update)
DE  15-DEC-1998 (Rel. 37, Last annotation update)
DE  INSULIN.
GN  INS.
OS  Anser anser anser (Western graylag goose), and
OC  Calina moschata (Muscovy duck).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Archosauria; Aves; Neognathae; Anseriformes; Anatidae; Anser.
OX  NCBI_TaxId=8844, 8855;
RN  [1]
RP  SEQUENCE.
RC  SPECIES-A. anser;
RA  Xu Y., Lin N., Zhang Y., Zhang Y.;
RT  "Isolation and sequence determination of goose insulin.";
RN  [2]
RP  SEQUENCE.
RC  SPECIES-C. moschata;
RX  MEDLINE=96321314; PubMed=8759296;
RA  Chevallier B., Anglade P., Derouet M., Mollé D., Simon J.;
RT  "Isolation and characterization of Muscovy (Calina moschata) duck insulin.";
RL  Comp. Biochem. Physiol. 114B:19-26(1996).
CC  -I- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC  -I- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
```

CC DISULFIDE BONDS.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
 CC PIR: JC0007: INGS.
 DR HSSP: P01308: INTS.
 DR InterPro: IPR000739: -.
 DR Pfam: PF00049: Insulin; 1.
 DR PRINTS: PR00276: INSULINA.
 DR PRINTS: PR00277: INSULINB.
 DR PROSITE: PS00262: INSULIN; 1.
 KW Insulin family; Hormone; Glucose metabolism.
 FT CHAIN 1 30 INSULIN B CHAIN.
 FT NON_CONS 30 31
 FT CHAIN 31 51 INSULIN A CHAIN.
 FT DISULFID 7 37 INTERCHAIN.
 FT DISULFID 19 50 INTERCHAIN.
 FT DISULFID 36 41
 SQ SEQUENCE 51 AA: 5716 MW: 976FEAED8C68386D CRC64:

Query Match 97.6%; Score 41; DB 1; Length 51;
 Best Local Similarity 100.0%; Pred. No. 0.13;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 GERGFY 8
 |||||
 Db 20 GERGFY 26

RESULT 5
 INS_BALBO STANDARD; PRT; 51 AA.
 AC P01314:
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE INSULIN.
 GN INS.
 OS Balanoptera borealis (Sei whale).
 OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;
 OC Balanopteridae; Balanoptera.
 OX NCBI_TaxId=9768;
 RN [1]
 RP SEQUENCE.
 RA Ishihara Y., Saito T., Ito Y., Fujino M.;
 RT "Structure of sperm- and sei-whale insulins and their breakdown by
 RT whale pepsin.";
 RL Nature 181:1468-1469(1958).
 CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
 CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
 CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
 CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
 CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
 CC DISULFIDE BONDS.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
 CC PIR: A01582: INMHIS.
 DR HSSP: P01315: 6INS.
 DR InterPro: IPR000739: -.
 DR Pfam: PF00049: Insulin; 1.
 DR PRINTS: PR00276: INSULINA.
 DR PRINTS: PR00277: INSULINB.
 DR PROSITE: PS00262: INSULIN; 1.
 KW Insulin family; Hormone; Glucose metabolism.
 FT CHAIN 1 30 INSULIN B CHAIN.
 FT NON_CONS 30 31
 FT CHAIN 31 51 INSULIN A CHAIN.
 FT DISULFID 7 37 INTERCHAIN.
 FT DISULFID 19 50 INTERCHAIN.
 FT DISULFID 36 41
 SQ SEQUENCE 51 AA: 5723 MW: 9007B50E400A7DDD CRC64:

Query Match 97.6%; Score 41; DB 1; Length 51;
 Best Local Similarity 100.0%; Pred. No. 0.13;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 GERGFY 8
 |||||
 Db 20 GERGFY 26

RESULT 6
 INS_BALPH STANDARD; PRT; 51 AA.
 AC P01312:
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE INSULIN.
 GN INS.
 OS Balanoptera physalus (Finback whale) (Common rorqual), and
 OS Physeter catodon (Sperm whale) (Physeter macrocephalus).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;
 OC Balanopteridae; Balanoptera.
 OX NCBI_TaxId=9770, 9735;
 RN [1]
 RP PARTIAL SEQUENCE.
 RC SPECIES-B.physalus;
 RA Hama H., Titani K., Sakaki S., Narita K.;
 RT "The amino acid sequence in fin-whale insulin.";
 RL J. Biochem. 56:285-293(1964).
 RN [2]
 RP SEQUENCE.
 RC SPECIES-P.catodon;
 RA Ishihara Y., Saito T., Ito Y., Fujino M.;
 RT "Structure of sperm- and sei-whale insulins and their breakdown by
 RT whale pepsin.";
 RL Nature 181:1468-1469(1958).
 RN [3]
 RP SEQUENCE.
 RC SPECIES-P.catodon;
 RA Harris J.I., Sanger F., Naughton M.A.;
 RT "Species differences in insulin.";
 RL Arch. Biochem. Biophys. 65:427-438(1956).
 CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
 CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
 CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
 CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
 CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
 CC DISULFIDE BONDS.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
 CC PIR: A01918: INMHP.
 DR PIR: A03142: INMHP.
 DR HSSP: P01315: 6INS.
 DR InterPro: IPR000739: -.
 DR Pfam: PF00049: Insulin; 1.
 DR PRINTS: PR00276: INSULINA.
 DR PRINTS: PR00277: INSULINB.
 DR PROSITE: PS00262: INSULIN; 1.
 KW Insulin family; Hormone; Glucose metabolism.
 FT CHAIN 1 30 INSULIN B CHAIN.
 FT NON_CONS 30 31
 FT CHAIN 31 51 INSULIN A CHAIN.
 FT DISULFID 7 37 INTERCHAIN.
 FT DISULFID 19 50 INTERCHAIN.
 FT DISULFID 36 41
 SQ SEQUENCE 51 AA: 5766 MW: 9007B514691A7CDD CRC64:

Query Match 97.6%; Score 41; DB 1; Length 51;
 Best Local Similarity 100.0%; Pred. No. 0.13;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY      2 GERGFY 8
RX      |||||||
DB      20 GERGFY 26

RESULT 7
INS_CAMDR
ID      INS_CAMDR      STANDARD;      PRT;      51 AA.
AC      P01320;
DT      21-JUL-1986 (Rel. 01, Created)
DT      21-JUL-1986 (Rel. 01, Last sequence update)
DE      01-FEB-1996 (Rel. 33, Last annotation update)
DE      INSULIN.
GN      INS.
OS      Camelus dromedarius (Dromedary) (Arabian camel).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.
OX      NCBI_TaxID=9838;
RN      (1)
RP      SEQUENCE.
RT      "The isolation and characterization of insulin of camel (Camelus
RT      dromedarius)."
RL      J. Fac. Med. Bagdad 14:16-28(1972).
CC      -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC      INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC      FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC      CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC      -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC      DISULFIDE BONDS.
CC      -1- SUBCELLULAR LOCATION: SECRETED.
CC      -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR      PIR: A02782; INCM.
DR      HSSP: P01317; 2INS.
DR      InterPro: IPR000739; -.
DR      Pfam: PF00049; Insulin; 1.
DR      PRINTS: PR00276; INSULIN.
DR      PRINTS: PR00277; INSULIN.
DR      PROSITE: PS00262; INSULIN; 1.
KW      Insulin family; Hormone; Glucose metabolism.
FT      CHAIN      1      30      INSULIN B CHAIN.
FT      NON_CONS      30      31
FT      CHAIN      31      51      INSULIN A CHAIN.
FT      DISULFID      7      37      INTERCHAIN.
FT      DISULFID      19      50      INTERCHAIN.
FT      DISULFID      36      41
SQ      SEQUENCE      51 AA; 5693 MW; 901E88BA085A7DDD CRC64;

Query Match      97.6%; Score 41; DB 1; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 GERGFY 8
RX      |||||||
DB      20 GERGFY 26

RESULT 8
INS_CAPHI
ID      INS_CAPHI      STANDARD;      PRT;      51 AA.
AC      P01319;
DT      21-JUL-1986 (Rel. 01, Created)
DT      21-JUL-1986 (Rel. 01, Last sequence update)
DE      01-FEB-1996 (Rel. 33, Last annotation update)
DE      INSULIN.
GN      INS.
OS      Capra hircus (Goat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC      Bovidae; Caprinae; Capra.
OX      NCBI_TaxID=9925;

```

```

RN      (1)
RP      SEQUENCE.
RX      MEDLINE=66160119; PubMed=5949593;
RA      Smith L.F.;
RT      "Species variation in the amino acid sequence of insulin.";
RL      Am. J. Med. 40:662-666(1966).
CC      -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC      INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC      FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC      CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC      -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC      DISULFIDE BONDS.
CC      -1- SUBCELLULAR LOCATION: SECRETED.
CC      -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR      PIR: A01586; INGT.
DR      HSSP: P01315; 9INS.
DR      InterPro: IPR000739; -.
DR      Pfam: PF00049; Insulin; 1.
DR      PRINTS: PR00276; INSULIN.
DR      PRINTS: PR00277; INSULIN.
DR      PROSITE: PS00262; INSULIN; 1.
KW      Insulin family; Hormone; Glucose metabolism.
FT      CHAIN      1      30      INSULIN B CHAIN.
FT      NON_CONS      30      31
FT      CHAIN      31      51      INSULIN A CHAIN.
FT      DISULFID      7      37      INTERCHAIN.
FT      DISULFID      19      50      INTERCHAIN.
FT      DISULFID      36      41
SQ      SEQUENCE      51 AA; 5692 MW; 9007B50CDB4E7DDD CRC64;

```

```

Query Match      97.6%; Score 41; DB 1; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      2 GERGFY 8
RX      |||||||
DB      20 GERGFY 26

RESULT 9
INS_DIDMA
ID      INS_DIDMA      STANDARD;      PRT;      51 AA.
AC      P18109;
DT      01-NOV-1990 (Rel. 16, Created)
DT      01-NOV-1990 (Rel. 16, Last sequence update)
DT      01-FEB-1996 (Rel. 33, Last annotation update)
DE      INSULIN.
DE      INS.
OS      Didelphis marsupialis virginiana (North American opossum).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Metatheria; Didelphimorphia; Didelphidae; Didelphis.
OX      NCBI_TaxID=9267;
RN      (1)
RP      SEQUENCE.
RC      TISSUE=Pancreas;
RX      MEDLINE=90160042; PubMed=2695899;
RA      Yu J.-H., Eng J., Rattan S., Yalow R.S.;
RT      "Opossum insulin, glucagon and pancreatic polypeptide: amino acid
RT      sequences."
RL      Peptides 10:1195-1197(1989).
CC      -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC      INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC      FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC      CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC      -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC      DISULFIDE BONDS.
CC      -1- SUBCELLULAR LOCATION: SECRETED.
CC      -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR      PIR: J00362; J00362.
DR      PIR: J00363; J00363.
DR      HSSP: P01317; 2INS.
DR      InterPro: IPR000739; -.

```


DR Pfam: PF00049; Insulin; 1.
 DR PRINTS: PR00276; INSULIN.
 DR PRINTS: PR00277; INSULIN.
 DR PROSITE: PS00262; INSULIN; 1.
 KW Insulin family; Hormone; Glucose metabolism.
 FT CHAIN 1 30 INSULIN B CHAIN.
 FT NON_CONS 30 31
 FT CHAIN 31 51 INSULIN A CHAIN.
 FT DISULFID 7 37 INTERCHAIN.
 FT DISULFID 19 50 INTERCHAIN.
 FT DISULFID 36 41
 SQ SEQUENCE 51 AA: 5732 MW: 9007B8BAE4BDEEDD CRC64;

Query Match 97.6%; Score 41; DB 1; Length 51;
 Best Local Similarity 100.0%; Pred. No. 0.13;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFY 8
 DB 20 GERGFY 26

RESULT 10
 INS_ELEMA STANDARD: PRT; 51 AA.
 AC P01316;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE INSULIN.
 GN INS.
 OS Elephas maximus (Indian elephant).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Proboscidea; Elephantidae; Elephas.
 OC NCBI_TaxID=9783;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=66160119; PubMed=5949593;
 RA Smith L.F.;
 RT *Species variation in the amino acid sequence of insulin.*;
 RU Am. J. Med. 40:662-666(1966).
 CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
 CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
 CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
 CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
 CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
 CC DISULFIDE BONDS.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- MISCELLANEOUS: THE SPECIES OF ELEPHANT IS NOT GIVEN, BUT IT IS
 CC MOST PROBABLY THE INDIAN ELEPHANT (ELEPHAS MAXIMUS).
 CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
 DR PIR: A01584; INCL.
 DR HSSP: P01308; IBERN.
 DR InterPro: IPR000739; .
 DR Pfam: PF00049; Insulin; 1.
 DR PRINTS: PR00276; INSULIN.
 DR PROSITE: PS00262; INSULIN; 1.
 KW Insulin family; Hormone; Glucose metabolism.
 FT CHAIN 1 30 INSULIN B CHAIN.
 FT NON_CONS 30 31
 FT CHAIN 31 51 INSULIN A CHAIN.
 FT DISULFID 7 37 INTERCHAIN.
 FT DISULFID 19 50 INTERCHAIN.
 FT DISULFID 36 41
 SQ SEQUENCE 51 AA: 5752 MW: 9007B50CDBA57D6D CRC64;

Query Match 97.6%; Score 41; DB 1; Length 51;
 Best Local Similarity 100.0%; Pred. No. 0.13;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFY 8
 DB 20 GERGFY 26

RESULT 11
 INS_FELCA STANDARD: PRT; 51 AA.
 AC P06306;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE INSULIN.
 GN INS.
 OS Felis silvestris catus (Cat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
 OC NCBI_TaxID=9685;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=86214076; PubMed=3518635;
 RA Hallden G., Gafvelin G., Mutt V., Joernvall H.;
 RT *Characterization of cat insulin.*;
 RL Arch. Biochem. Biophys. 247:20-27(1986).
 CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
 CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
 CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
 CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
 CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
 CC DISULFIDE BONDS.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
 DR PIR: A01588; INCL.
 DR HSSP: P01315; 9INS.
 DR InterPro: IPR000739; .
 DR Pfam: PF00049; Insulin; 1.
 DR PRINTS: PR00276; INSULIN.
 DR PROSITE: PS00262; INSULIN; 1.
 KW Insulin family; Hormone; Glucose metabolism.
 FT CHAIN 1 30 INSULIN B CHAIN.
 FT NON_CONS 30 31
 FT CHAIN 31 51 INSULIN A CHAIN.
 FT DISULFID 7 37 INTERCHAIN.
 FT DISULFID 19 50 INTERCHAIN.
 FT DISULFID 36 41
 SQ SEQUENCE 51 AA: 5745 MW: 9007B5096A0A7DDD CRC64;

Query Match 97.6%; Score 41; DB 1; Length 51;
 Best Local Similarity 100.0%; Pred. No. 0.13;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFY 8
 DB 20 GERGFY 26

RESULT 12
 INS_PLAFA STANDARD: PRT; 51 AA.
 AC P09477;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE INSULIN.
 GN INS.
 OS Platicthys flesus (European flounder).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
 OC Pleuronectoidei; Pleuronectidae; Platicthys.
 OC NCBI_TaxID=8260;

```

RN [1]
RP MEDLINE=87219793; PubMed=3556313;
RA Conlon J.M., Davis M.S., Thim L.;
RT "Primary structure of insulin and glucagon from the flounder
  (Platichthys flesus)."; 66:203-209(1987).
RL Gen. Comp. Endocrinol.
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
  INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
  FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
  CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
  DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC PIR: S07213; INFLE.
DR HSSP: P01315; 9INS.
DR InterPro: IPR000739; -.
DR Pfam: PF00049; INSULIN.
DR PRINTS: PR00277; INSULIN.
DR PROSITE: PS00262; INSULIN.
DR INSULIN family; Hormone; Glucose metabolism.
KM CHAIN 1 30 INSULIN B CHAIN.
FT CHAIN 31 31 INSULIN A CHAIN.
FT CHAIN 31 37 INTERCHAIN.
FT DISULFID 20 37 INTERCHAIN.
FT DISULFID 36 41 INTERCHAIN.
SQ SEQUENCE 51 AA; 5755 MW; 0AA232C873FF705B CRC64;

```

```

Query Match          97.6%; Score 41; DB 1; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2 GERGFY 8
   |||||
Db 21 GERGFY 27

```

```

RESULT 13
INS_TRASC STANDARD; PRT; 51 AA.
AC P3187;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE INSULIN.
GN INS.
OS Trachemys scripta (Red-eared slider turtle) (Pseudemys scripta), and
  Chrysemys dorsignif (Black-bellied slider turtle) (Trachemys dorsignif).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Testudines; Cryptodira; Testudinoidea; Emydidae; Trachemys.
OX NCBI_TaxID=34903, 31137;
RN [1]
RP SEQUENCE.
RC SPECIES=C.dorsignif; PubMed=1974347;
RC MEDLINE=90341082; PubMed=1974347;
RA Conlon J.M., Hicks J.W.;
RT "Isolation and structural characterization of insulin, glucagon and
  somatostatin from the turtle, Pseudemys scripta.";
RL Peptides 11:461-466(1990).
RN [2]
RP SEQUENCE.
RC SPECIES=C.dorsignif; PubMed=1808015;
RC MEDLINE=92225302; PubMed=1808015;
RA Cascone O., Turyn D., Dellacha J.M., Machado V.L.A., Marques M.,
  Vita N., Cassan C., Ferrara P., Guillemot J.C.;
RT "Isolation, purification and primary structure of insulin from the
  turtle Chrysemys dorsignif.";
RL Gen. Comp. Endocrinol. 84:355-359(1991).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT

```

```

CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
  DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC PIR: A60414; A60414.
DR PIR: A61129; A61129.
DR HSSP: P01308; 1HIS.
DR InterPro: IPR000739; -.
DR Pfam: PF00049; INSULIN.
DR PRINTS: PR00276; INSULIN.
DR PROSITE: PS00262; INSULIN.
DR INSULIN family; Hormone; Glucose metabolism.
KM CHAIN 1 30 INSULIN B CHAIN.
FT CHAIN 30 31 INSULIN A CHAIN.
FT CHAIN 31 37 INTERCHAIN (BY SIMILARITY).
FT DISULFID 7 37 INTERCHAIN (BY SIMILARITY).
FT DISULFID 19 50 INTERCHAIN (BY SIMILARITY).
FT DISULFID 36 41 BY SIMILARITY.
SQ SEQUENCE 51 AA; 5698 MW; 976EFA8987D40DD CRC64;

```

```

Query Match          97.6%; Score 41; DB 1; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2 GERGFY 8
   |||||
Db 20 GERGFY 26

```

```

RESULT 14
INS_ACTIGU STANDARD; PRT; 52 AA.
AC P81423;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE INSULIN.
GN INS.
OS Acipenser guldensadtii (Caspian sturgeon) (Russian sturgeon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Chondrostei; Acipenseriformes; Acipenseridae;
OC Acipenserinae; Acipenser.
OX NCBI_TaxID=7902;
RN [1]
RP SEQUENCE.
RC TISSUE=Pancreas; PubMed=9650713;
RC MEDLINE=98312733; PubMed=9650713;
RA Rusakov Y.I., Moriyama S., Bondareva V.M., Kolychev A.P., Amemiya Y.,
  Yasuda A., Kanauchi H.;
RT "Isolation and characterization of insulin in Russian sturgeon
  (Acipenser guldensadtii).";
J. Pept. Res. 51:395-400(1998).
RL -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
  INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
  FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
  CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
  DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR InterPro: IPR000739; -.
DR Pfam: PF00049; INSULIN.
DR PRINTS: PR00276; INSULIN.
DR PROSITE: PS00277; INSULIN.
DR PROSITE: PS00262; INSULIN.
DR INSULIN family; Hormone; Glucose metabolism.
KM CHAIN 1 31 INSULIN B CHAIN.
FT CHAIN 31 32 INSULIN A CHAIN.
FT NON_CONS 31 32

```

FT CHAIN 32 52 INSULIN A CHAIN.
 FT DISULFID 7 38 INTERCHAIN.
 FT DISULFID 19 51 INTERCHAIN.
 FT DISULFID 37 42
 SQ SEQUENCE 52 AA; 5811 MM; BOLD6939980D41631 CRC64;

Query Match 97.6%; Score 41; DB 1; Length 52;
 Best Local Similarity 100.0%; Pred. No. 0.14;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFFY 8
 DB 20 GERGFFY 26

RESULT 15
 INS_HYDCO
 ID INS_HYDCO STANDARD; PRT; 59 AA.
 AC P09536;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE INSULIN.
 GN INS.
 OS Hydrolaqus colliet (Spotted ratfish) (Pacific ratfish), and
 OS Chimera monstrosa (Rabbit fish).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
 OC Holocephali; Chimaeriformes; Chimaeridae; Hydrolaqus.
 OX NCBI_Taxid-7873, 7871.
 RN [1]
 RP SEQUENCE.
 RC SPECIES-H.colliet;
 RX MEDLINE-87054620; PubMed-3780981;
 RA Conlon J.M., Dafgard E., Falkner S., Thim L.;
 RT "The primary structure of ratfish insulin reveals an unusual mode of
 RT proinsulin processing.";
 RL FEBS Lett. 208:445-450(1986).
 RN [2]
 RP SEQUENCE.
 RC SPECIES-H.colliet;
 RX MEDLINE-89133911; PubMed-2646172;
 RA Conlon J.M., Goelke R., Andrews P.C., Thim L.;
 RT "Multiple molecular forms of insulin and glucagon-like peptide from
 RT the Pacific ratfish (Hydrolaqus colliet).";
 RL Gen. Comp. Endocrinol. 73:136-146(1989).
 RN [3]
 RP SEQUENCE.
 RC SPECIES-C.monstrosa;
 RX MEDLINE-89031910; PubMed-3053327;
 RA Conlon J.M., Andrews P.C., Falkner S., Thim L.;
 RT "Isolation and structural characterization of insulin from the
 RT holocephalan fish, Chimera monstrosa (rabbit fish).";
 RL Gen. Comp. Endocrinol. 72:154-160(1988).
 CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
 CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
 CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
 CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
 CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
 CC DISULFIDE BONDS.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- MISCELLANEOUS: DUE TO A SUBSTITUTION OF THE ARG IN POSITION 31 BY
 CC AN ILE, THIS INSULIN B CHAIN IS LONGER THEN MOST OTHER B CHAINS
 CC AND IS PROCESSED DIFFERENTLY.
 CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
 DR PIR: S06474; INFI.
 DR PIR: S06420; INRO.
 DR PIR: A24634; A24634.
 DR HSSP: P01308; IHIS.
 DR InterPro: IPR000739; -
 DR Pfam: PF00049; INSULIN.1.
 DR PRINTS: PRO0276; INSULIN.A.
 DR PRINTS: PRO0277; INSULIN.B.

DR PROSITE: P500262; INSULIN.1.
 KW Insulin family; Hormone; Glucose metabolism.
 FT CHAIN 1 38 INSULIN B CHAIN.
 FT NON_CONS 38 39
 FT CHAIN 39 59 INSULIN A CHAIN.
 FT DISULFID 7 45 INTERCHAIN.
 FT DISULFID 19 58 INTERCHAIN.
 FT DISULFID 44 49
 SQ SEQUENCE 59 AA; 6606 MM; 8827A57A9ED6D4AC CRC64;

Query Match 97.6%; Score 41; DB 1; Length 59;
 Best Local Similarity 100.0%; Pred. No. 0.15;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFFY 8
 DB 20 GERGFFY 26

Search completed: September 6, 2001, 16:51:08
 Job time: 812 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:39:37 : Search time 113.12 seconds
(without alignments)
1.456 Million cell updates/sec

Title: US-09-603-713-23

Perfect score: 42

Sequence: 1 XGERGFY 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*

1: /cgn2_6/prodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/prodata/2/1aa/Packfillsl.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	97.6	12	6 5227466-1	Patent No. 5227466
2	41	97.6	27	1 US-08-304-070-3	Sequence 3, Appl1
3	41	97.6	29	1 US-08-304-070-4	Sequence 4, Appl1
4	41	97.6	29	1 US-08-304-070-2	Sequence 2, Appl1
5	41	97.6	29	2 US-08-992-676-3	Sequence 3, Appl1
6	41	97.6	29	6 5164366-4	Patent No. 5164366
7	41	97.6	29	6 5514646-49	Patent No. 5514646
8	41	97.6	30	1 US-08-212-696-2	Sequence 2, Appl1
9	41	97.6	30	1 US-08-233-617-2	Sequence 2, Appl1
10	41	97.6	30	1 US-08-160-376A-2	Sequence 2, Appl1
11	41	97.6	30	1 US-08-304-070-2	Sequence 2, Appl1
12	41	97.6	30	1 US-08-301-838-2	Sequence 2, Appl1
13	41	97.6	30	1 US-08-389-487-5	Sequence 5, Appl1
14	41	97.6	30	1 US-08-342-931-2	Sequence 2, Appl1
15	41	97.6	30	1 US-08-400-256-2	Sequence 2, Appl1
16	41	97.6	30	2 US-08-508-664-7	Sequence 7, Appl1
17	41	97.6	30	2 US-08-353-476-86	Sequence 8, Appl1
18	41	97.6	30	2 US-08-484-219-6	Sequence 6, Appl1
19	41	97.6	30	2 US-08-979-587-2	Sequence 2, Appl1
20	41	97.6	30	2 US-08-992-676-2	Sequence 2, Appl1
21	41	97.6	30	2 US-09-134-836-2	Sequence 2, Appl1
22	41	97.6	30	3 US-08-975-365-2	Sequence 2, Appl1
23	41	97.6	30	3 US-08-750-391-4	Sequence 2, Appl1
24	41	97.6	30	3 US-08-622-046B-20	Sequence 20, Appl1
25	41	97.6	30	3 US-09-255-668-1	Sequence 1, Appl1
26	41	97.6	30	4 US-09-099-307-2	Sequence 2, Appl1
27	41	97.6	30	4 US-09-099-307-2	Sequence 2, Appl1

28	41	97.6	30	4 US-09-099-307-3	Sequence 3, Appl1
29	41	97.6	30	4 US-09-099-307-4	Sequence 4, Appl1
30	41	97.6	30	4 US-09-099-307-5	Sequence 5, Appl1
31	41	97.6	30	4 US-09-099-307-10	Sequence 10, Appl1
32	41	97.6	30	4 US-08-800-574-2	Sequence 2, Appl1
33	41	97.6	30	5 PCT-US94-04179-3	Sequence 3, Appl1
34	41	97.6	30	5 PCT-US94-04179-4	Sequence 4, Appl1
35	41	97.6	30	5 PCT-US94-04179-5	Sequence 5, Appl1
36	41	97.6	30	6 5514646-51	Patent No. 5514646
37	41	97.6	30	6 5514646-52	Patent No. 5514646
38	41	97.6	31	1 US-08-389-487-10	Sequence 10, Appl1
39	41	97.6	31	2 US-08-992-676-5	Sequence 5, Appl1
40	41	97.6	32	1 US-08-087-831-1	Sequence 1, Appl1
41	41	97.6	32	1 US-08-087-831-2	Sequence 2, Appl1
42	41	97.6	32	1 US-08-389-487-9	Sequence 9, Appl1
43	41	97.6	32	1 US-08-466-945-1	Sequence 1, Appl1
44	41	97.6	32	1 US-08-466-945-2	Sequence 2, Appl1
45	41	97.6	32	2 US-08-992-676-4	Sequence 4, Appl1

ALIGNMENTS

RESULT 1
5227466-1
; Patent No. 5227466
; APPLICANT: DEMEYTS, PIERRE
; TITLE OF INVENTION: INSULIN RECEPTOR BINDING SITE
; NUMBER OF SEQUENCES: 32
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/213,918
; FILING DATE: 30-JUN-1988
; SEQ ID NO: 1:
; LENGTH: 12
5227466-1

Query Match 97.6%; Score 41; DB 6; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.043;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GERGFY 8
Db 2 GERGFY 8

RESULT 2
US-08-304-070-3
; Sequence 3, Application US/08304070
; Patent No. 5547929
; GENERAL INFORMATION:
; APPLICANT: Anderson Jr., James H.
; APPLICANT: De Felippis, Michael R.
; APPLICANT: Frank, Bruce H.
; APPLICANT: Havel, Henry A.
; TITLE OF INVENTION: INSULIN ANALOG FORMULATIONS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company/Patent Division
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/304,070
; FILING DATE: 12-SEP-1994
; CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:
NAME: Caltrider, Steven P.
REGISTRATION NUMBER: 36,467
REFERENCE/DOCKET NUMBER: X-9635
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-0757
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 27 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-304-070-3

Query Match 97.6%; Score 41; DB 1; Length 27;
Best Local Similarity 100.0%; Pred. No. 0.095; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0;

OY 2 GERGFY 8
DB 20 GERGFY 26

RESULT 3
US-08-304-070-4
Sequence 4, Application US/08304070
Patent No. 5547929
GENERAL INFORMATION:
APPLICANT: Anderson Jr., James H.
APPLICANT: De Felippis, Michael R.
APPLICANT: Frank, Bruce H.
APPLICANT: Havel, Henry A.
TITLE OF INVENTION: INSULIN ANALOG FORMULATIONS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company/Patent Division
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana
COUNTRY: United States
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/304,070
FILING DATE: 12-SEP-1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Caltrider, Steven P.
REGISTRATION NUMBER: 36,467
REFERENCE/DOCKET NUMBER: X-9635
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-0757
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 29 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-304-070-4

Query Match 97.6%; Score 41; DB 1; Length 29;
Best Local Similarity 100.0%; Pred. No. 0.1; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0;

OY 2 GERGFY 8

DB 20 GERGFY 26

RESULT 4
US-08-430-273-2
Sequence 2, Application US/08430273
Patent No. 5621073
GENERAL INFORMATION:
APPLICANT: Dickhardt, Rainer
APPLICANT: Unger, Bernhard
APPLICANT: Grate, Claudia
TITLE OF INVENTION: Process for Obtaining Insulin-Containing
TITLE OF INVENTION: Solutions
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farbow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/430,273
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/991,261
FILING DATE: 16-DEC-1992
APPLICATION NUMBER: DE P 4141794.1
FILING DATE: 18-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE P 4220293.0
FILING DATE: 20-JUN-1992
ATTORNEY/AGENT INFORMATION:
NAME: Finaudi, Carol P.
REGISTRATION NUMBER: 32,220
REFERENCE/DOCKET NUMBER: 02481-1242-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 29 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-430-273-2

Query Match 97.6%; Score 41; DB 1; Length 29;
Best Local Similarity 100.0%; Pred. No. 0.1; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0;

OY 2 GERGFY 8
DB 20 GERGFY 26

RESULT 5
US-08-992-676-3
Sequence 3, Application US/08992676
Patent No. 5977297
GENERAL INFORMATION:
APPLICANT: OBERMEIER, Rainer
APPLICANT: LUDWIG, Jürgen
APPLICANT: SABEL, Walter

TITLE OF INVENTION: A Process for Isolating Insulin Using
TITLE OF INVENTION: High-Pressure Liquid Chromatography
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: FINNEGAN HENDERSON FARABOW GARRETT & DUNNER,
ADDRESS: L.L.P.
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/992.676
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Jensen, Allen R.
REGISTRATION NUMBER: 28,224
REFERENCE/DOCKET NUMBER: 02481.1567-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 29 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
US-08-992-676-3

Query Match 97.6%; Score 41; DB 2; Length 29;
Best Local Similarity 100.0%; Pred. No. 0.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFFY 8
DB 20 GERGFFY 26

RESULT 6
5164366-4
PATENT NO. 5164366
APPLICANT: BALSCHMIDT, PER, BRANGE, JENS J.V.
TITLE OF INVENTION: HUMAN INSULIN ANALOGUES
NUMBER OF SEQUENCES: 14
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/453,445
FILING DATE: 20-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 416,218
FILING DATE: 2-OCT-1989
APPLICATION NUMBER: 332,697
FILING DATE: 03-APR-1989
SEQ ID NO: 4:
LENGTH: 29
5164366-4

Query Match 97.6%; Score 41; DB 6; Length 29;
Best Local Similarity 100.0%; Pred. No. 0.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFFY 8
DB 20 GERGFFY 26

DB 20 GERGFFY 26
RESULT 7
514646-49
PATENT NO. 514646
APPLICANT: CHANCE, RONALD E.; DIMARCHI, RICHARD D.; FRANK,
BRUCE H.; SHIELDS, JAMES E.
TITLE OF INVENTION: INSULIN ANALOGS MODIFIED AT POSITION
29 OF THE B CHAIN
NUMBER OF SEQUENCES: 52
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/57,201
FILING DATE: 05-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 686,632
FILING DATE: 17-APR-1991
APPLICATION NUMBER: 388,201
FILING DATE: 04-AUG-1989
APPLICATION NUMBER: 308,352
FILING DATE: 09-FEB-1989
SEQ ID NO: 49:
LENGTH: 29
514646-49

Query Match 97.6%; Score 41; DB 6; Length 29;
Best Local Similarity 100.0%; Pred. No. 0.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFFY 8
DB 19 GERGFFY 25

RESULT 8
US-08-212-696-2
SEQUENCE 2, Application US/08212696
PATENT NO. 5422339
GENERAL INFORMATION:
APPLICANT: George S. Eisenbarth et al.
TITLE OF INVENTION: PEPTIDES HAVING INSULIN
TITLE OF INVENTION: AUTOANTIBODY BUT NOT
TITLE OF INVENTION: INSULIN RECEPTOR BINDING
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
COMPUTER: IBM PS/2 Model 502 or 555X
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: Wordperfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/212,696
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/671,455
FILING DATE: 03/19/91
ATTORNEY/AGENT INFORMATION:
NAME: Freeman, John W.
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 00303/009001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

LENGTH: 30
TYPE: amino acid
STRANDEDNESS: N/A
TOPOLOGY: linear
US-08-212-696-2

Query Match 97.6%; Score 41; DB 1; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFY 8
DB 20 GERGFY 26

RESULT 9

US-08-233-617-2
Sequence 2, Application US/08233617
Patent No. 5466666
GENERAL INFORMATION:
APPLICANT: Obermeyer, Rainer
APPLICANT: Sabell, Walter
APPLICANT: Geisen, Karl
TITLE OF INVENTION: Amorphous Monospherical Forms of Insulin
TITLE OF INVENTION: Derivatives
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-1315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,617
FILING DATE: 25-APR-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: P 43 13 702.4
FILING DATE: 27-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Carol P. Elinaudi
REGISTRATION NUMBER: 32,220
REFERENCE/DOCKET NUMBER: 02481.1374-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
US-08-233-617-2

DB 20 GERGFY 26

RESULT 10

US-08-160-376A-2
Sequence 2, Application US/08160376A
Patent No. 5473049
GENERAL INFORMATION:
APPLICANT: Obermeyer, Rainer
APPLICANT: Gerl, Martin
APPLICANT: Ludwig, Jurgen
APPLICANT: Sabell, Walter
TITLE OF INVENTION: Process For Obtaining Proinsulin
TITLE OF INVENTION: Possessing Correctly Linked
CYCLINE BRIDGES
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kenneth A. Genoni, Esq.
STREET: Rt. 202-206 No. 5473049th/P.O. Box 2500
CITY: Somerville
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 08876-1258
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
COMPUTER: IBM 386
OPERATING SYSTEM: WINDOWS 3.1
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/160,376A
FILING DATE: December 1, 1993
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GE P 4240420.7
FILING DATE: December 2, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Baird V. Maurer, Esq.
REGISTRATION NUMBER: 31,287
REFERENCE/DOCKET NUMBER: HOE 92/F 384
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 231-4079
TELEFAX: (908) 231-2255
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 Amino Acids
TYPE: Amino Acid
TOPOLOGY: not relevant
US-08-160-376A-2

Query Match 97.6%; Score 41; DB 1; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFY 8
DB 20 GERGFY 26

RESULT 11

US-08-304-070-2
Sequence 2, Application US/08304070
Patent No. 5547929
GENERAL INFORMATION:
APPLICANT: Anderson Jr., James H.
APPLICANT: De Felippis, Michael R.
APPLICANT: Frank, Bruce H.
APPLICANT: Havel, Henry A.
TITLE OF INVENTION: INSULIN ANALOG FORMULATIONS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company/Patent Division

OY 2 GERGFY 8

Query Match 97.6%; Score 41; DB 1; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana
COUNTRY: United States
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/304,070
FILING DATE: 12-SEP-1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Caltrider, Steven P.
REGISTRATION NUMBER: 36,467
REFERENCE/DOCKET NUMBER: X-9635
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-0757
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 28
OTHER INFORMATION: /note="Xaa at position 28 of SEQ
OTHER INFORMATION: ID NO:2 is Asp, Lys, Leu, Val, or Ala."
FEATURE:
NAME/KEY: Modified-site
LOCATION: 29
OTHER INFORMATION: /note="Xaa at position 29 of SEQ
OTHER INFORMATION: ID NO:2 is Lys or Pro."
US-08-304-070-2

Query Match 97.6%; Score 41; DB 1; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFY 8
Db 20 GERGFY 26

RESULT 12
US-08-301-838-2
Sequence 2, Application US/08301838
Patent No. 5597796
GENERAL INFORMATION:
APPLICANT: Brange, Jens J.V.
TITLE OF INVENTION: TRANSDERMAL INSULIN
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: NO. 55977960 No. 5597796disk of No. 5597796th America, Inc.
STREET: 405 Lexington Avenue, 64th floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/301,838
FILING DATE:
CLASSIFICATION: 514

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/085,473
FILING DATE: 30-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/751,836
FILING DATE: 30-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DK 0101/91
FILING DATE: 22-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/DK92/00019
FILING DATE: 22-JAN-1992
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3447.210-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 7
OTHER INFORMATION: /note="This disulfide bond is
OTHER INFORMATION: between residue number 7 of this sequence and
OTHER INFORMATION: residue number 7 of SEQ ID NO:1"
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 19
OTHER INFORMATION: /note="This disulfide bond is
OTHER INFORMATION: between residue number 19 of this sequence and
OTHER INFORMATION: residue number 20 of SEQ ID NO:1"
US-08-301-838-2

Query Match 97.6%; Score 41; DB 1; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFY 8
Db 20 GERGFY 26

RESULT 13
US-08-389-487-5
Sequence 5, Application US/08389487
Patent No. 5663291
GENERAL INFORMATION:
APPLICANT: Obermeier, Rainer
APPLICANT: Gerl, Martin
APPLICANT: Ludwig, Jurgen
APPLICANT: Sabel, Walter
TITLE OF INVENTION: Process for Obtaining Insulin Having
TITLE OF INVENTION: Correctly Linked Cysteine Bridges
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/389,487
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Einaudi, Carol P.
REGISTRATION NUMBER: 32,220
REFERENCE/DOCKET NUMBER: 02481.1424-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4000
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-389-487-5

Query Match 97.6%; Score 41; DB 1; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GERGFFY 8
DB 20 GERGFFY 26

RESULT 14

US-08-342-931-2
Sequence 2, Application US/08342931
Patent No. 5693609
GENERAL INFORMATION:
APPLICANT: Baker et. al.
TITLE OF INVENTION: Acylated Insulin Analogs
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/342,931
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Steven P. Caltrider
REGISTRATION NUMBER: 36467
REFERENCE/DOCKET NUMBER: X9720
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317) 276-0757
TELEFAX: (317) 277-1917
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: polypeptide
FEATURE:
NAME/KEY: Variable Site
LOCATION: 1
IDENTIFICATION METHOD:
OTHER INFORMATION: "Xaa at position 1 of SEQ ID NO:2 is Phe; or acylated Phe
FEATURE:
NAME/KEY: Variable Site
LOCATION: 28
IDENTIFICATION METHOD:
OTHER INFORMATION: "Xaa at position 28 of SEQ ID NO:2 is Asp, Lys, Leu, Val,
FEATURE:
NAME/KEY: Variable Site
LOCATION: 29
IDENTIFICATION METHOD:
OTHER INFORMATION: "Xaa at position 29 of SEQ ID NO:2 is Lys, Pro; or acylate
US-08-342-931-2

Query Match 97.6%; Score 41; DB 1; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GERGFFY 8
DB 20 GERGFFY 26

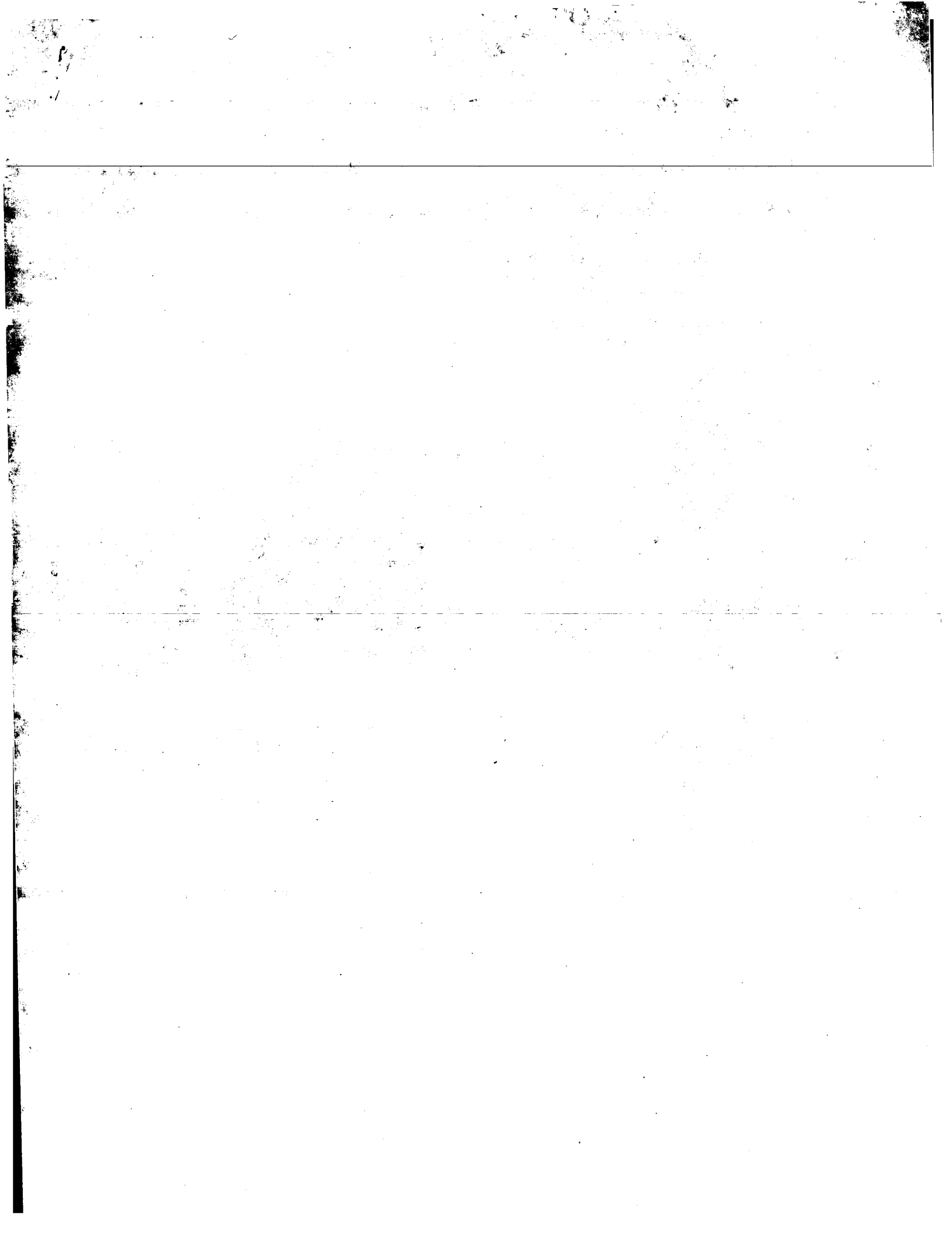
RESULT 15

US-08-400-256-2
Sequence 2, Application US/08400256
Patent No. 5750497
GENERAL INFORMATION:
APPLICANT: Havelund, Svend
APPLICANT: Halstrom, John
APPLICANT: Jonassen, Ib
APPLICANT: Andersen, Jan
APPLICANT: Markussen, Jan
TITLE OF INVENTION: ACYLATED INSULIN
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 57504970 No. 5750497disk of No. 5750497th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/400,256
FILING DATE: 03-MAR-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3985.220-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-878-9655
TELEFAX: 212-867-0123
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-400-256-2

Query Match 97.6%; Score 41; DB 1; Length 30;

Best Local Similarity 100.0%; Pred. No. 0.11;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 2 GERGFY 8
 111111
 Db 20 GERGFY 26

Search completed: September 6, 2001, 16:39:37
 Job time: 126 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:43:29 ; Search time 225.25 Seconds
(without alignments)
2.153 Million cell updates/sec

Title: US-09-603-713-23

Sequence: 1 XGERGFFY 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

Database : A_Geneseq_0601:*

1:	/S1D8/gcgdata/geneseq/geneseq/AA1960.DAT *
2:	/S1D8/gcgdata/geneseq/geneseq/AA1961.DAT *
3:	/S1D8/gcgdata/geneseq/geneseq/AA1962.DAT *
4:	/S1D8/gcgdata/geneseq/geneseq/AA1963.DAT *
5:	/S1D8/gcgdata/geneseq/geneseq/AA1964.DAT *
6:	/S1D8/gcgdata/geneseq/geneseq/AA1965.DAT *
7:	/S1D8/gcgdata/geneseq/geneseq/AA1966.DAT *
8:	/S1D8/gcgdata/geneseq/geneseq/AA1967.DAT *
9:	/S1D8/gcgdata/geneseq/geneseq/AA1968.DAT *
10:	/S1D8/gcgdata/geneseq/geneseq/AA1969.DAT *
11:	/S1D8/gcgdata/geneseq/geneseq/AA1970.DAT *
12:	/S1D8/gcgdata/geneseq/geneseq/AA1971.DAT *
13:	/S1D8/gcgdata/geneseq/geneseq/AA1972.DAT *
14:	/S1D8/gcgdata/geneseq/geneseq/AA1973.DAT *
15:	/S1D8/gcgdata/geneseq/geneseq/AA1974.DAT *
16:	/S1D8/gcgdata/geneseq/geneseq/AA1975.DAT *
17:	/S1D8/gcgdata/geneseq/geneseq/AA1976.DAT *
18:	/S1D8/gcgdata/geneseq/geneseq/AA1977.DAT *
19:	/S1D8/gcgdata/geneseq/geneseq/AA1978.DAT *
20:	/S1D8/gcgdata/geneseq/geneseq/AA1979.DAT *
21:	/S1D8/gcgdata/geneseq/geneseq/AA2000.DAT *
22:	/S1D8/gcgdata/geneseq/geneseq/AA2001.DAT *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	41	97.6	8	18	AAW29334	Peptide fragment h
2	41	97.6	8	22	AA86581	Oxidised Insulin B
3	41	97.6	8	22	AA861343	Mempsin 2 substra
4	41	97.6	12	11	AA803176	Residues 19-30 of
5	41	97.6	12	19	AAW59281	Homo sapiens insuli
6	41	97.6	20	21	AAV99233	H1A class II bindi
7	41	97.6	26	22	AA848863	Mutant human insuli
8	41	97.6	27	13	AA829997	Human insulin B-ct
9	41	97.6	27	13	AA830006	Human insulin B-ct
10	41	97.6	27	13	AA830007	Human insulin B-ct
11	41	97.6	27	13	AA830008	Human insulin B-ct

12	41	97.6	27	13	AAAR0009	Human Insulin B-chain
13	41	97.6	28	19	AAAR6931	Human Insulin B-chain
14	41	97.6	28	19	AAAR6932	Human Insulin B-chain
15	41	97.6	28	19	AAAR6924	Human Insulin B-chain
16	41	97.6	28	19	AAAR6925	Human Insulin B-chain
17	41	97.6	29	3	AAAP20139	Sequence of des-ph
18	41	97.6	29	3	AAAP20262	Modified Insulin B
19	41	97.6	29	6	AAAP50834	Sequence of des-ph
20	41	97.6	29	11	AAAR08187	Insulin derivative
21	41	97.6	29	11	AAAR08467	Example of pref. m
22	41	97.6	29	15	AAAR52573	Insulin B chain fr
23	41	97.6	29	19	AAAR53756	Insulin protein fr
24	41	97.6	29	19	AAAR60061	Human Insulin B-cp
25	41	97.6	29	19	AAAR60062	Human Insulin B-cp
26	41	97.6	29	19	AAAR6928	Human Insulin B-cp
27	41	97.6	29	19	AAAR6930	Human Insulin B-cp
28	41	97.6	29	19	AAAR6935	Human Insulin B-cp
29	41	97.6	29	19	AAAR1337	Human Insulin B-cp
30	41	97.6	29	19	AAAR6922	Insulin B chain an
31	41	97.6	29	19	AAAR6923	Insulin B chain an
32	41	97.6	29	19	AAAR6926	Insulin B chain an
33	41	97.6	30	3	AAAP20316	Insulin B chain de
34	41	97.6	30	3	AAAP20263	Insulin chain B an
35	41	97.6	30	6	AAAP50835	Sequence of des-Th
36	41	97.6	30	6	AAAP50826	Sequence of human
37	41	97.6	30	6	AAAP51010	Sequence of human
38	41	97.6	30	7	AAAP60909	Derivative of Insu
39	41	97.6	30	7	AAAP60911	Derivative of Insu
40	41	97.6	30	7	AAAP61329	Modified Insulin B
41	41	97.6	30	7	AAAP61330	Modified Insulin B
42	41	97.6	30	7	AAAP61333	Modified Insulin B
43	41	97.6	30	7	AAAP61335	Modified Insulin B
44	41	97.6	30	8	AAAP70896	Sequence of B-chain
45	41	97.6	30	8	AAAP71023	Sequence of B-chain

ALIGNMENTS

RESULT

ID	AAW29334	standard; peptide; 8 AA.

AC AAW29334;

DT 21-APR-1998 (first entry)

DE Peptide fragment having insulin activity useful for treating diabetes.

KW Insulin activity; diabetes; insulin receptor; hormone; disulphide.

Synthetic

FH	Key	Location/Qualifiers
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

/note= "This Cys is disulphide bonded to Cys-Asn"

FT /note= "This amino acid is incorrectly shown as Gly

PN RU2078769-C1.

PD 10-MAY-1997.

PF 18-AUG-1995; 95RU-0114858.

PR 18-AUG-1995; 95RU-0114858.

PA (AMBI-) A MED BIOMED CHEM RES INST.

PI Archkaov AI, Dyumaev KM, Knyazhev VA,

DR WPI; 1997-510880/47.

XX New peptide fragment having insulin activity - useful for treating
 PT diabetes or for studying insulin receptor function
 PS
 XX Claim 1: Columns 7-8: 5pp; Russian.
 XX
 CC This sequence represents a peptide fragment with biological activity
 CC comparable to that of native insulin hormone and will be useful in a
 CC biochemistry and medicine. For example, the peptide can be used in a
 CC medicinal preparation for treatment of insulin diabetes and in studies
 CC of molecular hormone-receptor mechanism.
 CC
 SQ Sequence 8 AA;
 QY
 Query Match 97.6%; Score 41; DB 18; Length 8;
 Best Local Similarity 100.0%; Pred. No. 3.4e+05; Indels 0; Gaps 0;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 2 GERGFY 8
 2 gergify 8
 Db 2 gergify 8
 RESULT 2
 AAB6581
 ID AAB6581 standard; Peptide; 8 AA.
 XX
 AC AAB6581;
 XX
 DT 12-APR-2001 (first entry)
 XX
 DE Oxidised insulin B-chain peptide #2.
 XX
 KW memapsin 2; neuroprotective; amyloid precursor protein;
 KW APP; memapsin 2 inhibitor; Alzheimer's disease; insulin B-chain peptide.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Modified-site 1
 FT /note= "cysteic acid"
 XX
 FN WO200100665-A2.
 PN
 XX
 PD 04-JAN-2001.
 XX
 PE 27-JUN-2000; 2000WO-US17742.
 XX
 PR 28-JUN-1999; 99US-0141363.
 PR 30-NOV-1999; 99US-0168060.
 PR 25-JAN-2000; 2000US-0177836.
 PR 27-JAN-2000; 2000US-0178368.
 PR 08-JUN-2000; 2000US-0210292.
 XX
 PA (OKLA-) OKLAHOMA MEDICAL RES FOUND.
 PA (UNIT) UNIT ILLINOIS FOUND.
 XX
 PI Tang JUN, Hong L, Ghosh AK;
 XX
 DR WPI; 2001-137933/14.
 XX
 PT Novel memapsin 2 inhibitors which bind to active site of memapsin 2
 PT having 2 catalytic aspartic residues and substrate binding cleft, used
 PT to treat Alzheimer's disease by blocking amyloid precursor protein
 PT cleavage -
 PS
 XX Example 4; Page 33; 86pp; English.
 CC The present sequence is given in a specification relating to an inhibitor
 CC of catalytically active memapsin 2. The inhibitor binds to the memapsin 2
 CC active site, which is defined by the presence of two catalytic aspartic
 CC residues and a substrate binding cleft. The inhibitor is useful for

CC the treatment and diagnosis of Alzheimer's disease. It is useful in
 CC screens for individuals with a genetic predisposition to Alzheimer's
 CC disease. The inhibitor is useful as a reagent for specifically binding to
 CC memapsin 2 or memapsin 2 analogues and for aiding in memapsin 2
 CC isolation, purification and characterisation.
 CC
 SQ Sequence 8 AA;
 QY
 Query Match 97.6%; Score 41; DB 22; Length 8;
 Best Local Similarity 100.0%; Pred. No. 3.4e+05; Indels 0; Gaps 0;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 2 GERGFY 8
 2 gergify 8
 Db 2 gergify 8
 RESULT 3
 AAB61343
 ID AAB61343 standard; peptide; 8 AA.
 XX
 AC AAB61343;
 XX
 DT 02-APR-2001 (first entry)
 XX
 DE Memapsin 2 substrate #2.
 XX
 KW Memapsin 2; catalyst; Alzheimer's.
 XX
 OS unidentified.
 XX
 PN WO200100663-A2.
 PD
 PD 04-JAN-2001.
 XX
 PF 27-JUN-2000; 2000WO-US17661.
 XX
 PR 28-JUN-1999; 99US-0141363.
 PR 30-NOV-1999; 99US-0168060.
 PR 25-JAN-2000; 2000US-0177836.
 PR 27-JAN-2000; 2000US-0178368.
 PR 08-JUN-2000; 2000US-0210292.
 XX
 PA (OKLA-) OKLAHOMA MEDICAL RES FOUND.
 PA
 XX
 PI Tang JUN, Lin X, Koelsch G;
 XX
 DR WPI; 2001-102885/11.
 XX
 PT Purified recombinant catalytically active memapsin 2, used to screen
 PT inhibitors of it, which are used to treat and prevent Alzheimer's
 PT disease -
 PS
 XX Example 4; Page 33; 86pp; English.
 CC The present invention relates to a purified recombinant
 CC catalytically active memapsin 2. The invention may be used for
 CC isolating inhibitors which are used to treat or prevent
 CC Alzheimer's disease. The invention may also be used to screen
 CC for individuals more genetically prone to develop Alzheimer's
 CC disease.
 CC
 SQ Sequence 8 AA;
 QY
 Query Match 97.6%; Score 41; DB 22; Length 8;
 Best Local Similarity 100.0%; Pred. No. 3.4e+05; Indels 0; Gaps 0;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 2 GERGFY 8
 2 gergify 8
 Db 2 gergify 8

```

RESULT 4
AA03176 ID AAR03176 standard; protein; 12 AA.
XX
AC AAR03176;
XX
DT 04-AUG-1990 (first entry)
XX
DE Residues 19-30 of C-terminal end of insulin B-chain.
XX
KW Human insulin B-chain.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 1..2 /note="Invariant or mostly invariant insulin residues"
FT Region 22..25 /note="as above"
FT Region 26 /note="as above"
FT Region /note="as above"
XX
PN WO9000562-A.
XX
PD 25-JAN-1990.
XX
PF 30-JUN-1989; 89WO-US02830.
XX
PR 30-DEC-1988; 88US-0292099, US-213918.
XX
PA (DEME ) DEMEYTS P.
XX
PI Demeys P;
XX
DR WPI; 1990-051687/07.
XX
PT Insulinomimetic and insulin receptor binding site peptide(s) -
PT used for treating diabetes and for studying insulin mechanism
XX
PS Disclosure; 38pp; English.
XX
CC It is the C-terminal end of the insulin B-chain. It is compared in the
CC patent with various receptor fragments, containing at least two aromatic
CC side chains, to show homology. The numbering of the bases is according to
CC Ulrich's numbering.
XX
SQ Sequence 12 AA;

Query Match 97.6%; Score 41; DB 11; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFY 8
DB 2 gergffy 8

RESULT 5
AA059281 ID AAM59281 standard; Protein; 12 AA.
XX
AC AAM59281;
XX
DT 11-SEP-1998 (first entry)
XX
DE Homo sapiens insulin B urokinase cleavage site.
XX
KW Insulin B; proproteins; protease-activatable; cancer; specific;
KW selective; treatment; urokinase; cleavage site.
XX

```

```

OS Homo sapiens.
XX
PN WO9820135-A2.
XX
PD 14-MAY-1998.
XX
PF 05-NOV-1997; 97WO-US20207.
XX
PR 06-NOV-1996; 96US-0030376.
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Fitzgerald DJ, Pastan I, Rafter Y;
XX
DR WPI; 1998-286951/25.
XX
PT Pseudomonas exotoxin A-like proprotein which is protease-activatable
PT - allows activation by desired protease through protease activatable
PT sequence in domain II loop, useful to selectively kill e.g. cancer
PT cells
XX
PS Disclosure; Page 29; 74pp; English.
XX
CC The sequence is that of an insulin B urokinase cleavage
CC site which was used in construction of a protease-activatable
CC Pseudomonas exotoxin (PE) A-like proprotein. Such proproteins
CC can be used to kill (especially prostate or colon) cancer
CC cells. They are modified for activation by a desired protease
CC by insertion of a protease activatable sequence in the domain
CC II loop; proprotein activation results in formation of cytotoxic
CC PE. PE is normally translocated into the cytosol after cleavage of a
CC furin recognition site in domain II by furin, but in the proproteins the
CC furin recognition site is replaced by a site recognised by a protease
CC made/secreted by a cell targeted for death, e.g. a cancer cell. The
CC proproteins can be used in vivo e.g. to treat mammals suffering from
CC cancer or ex vivo e.g. to selectively eliminate cultured
CC mammalian cells prior to reintroducton. Mammalian cells can be
CC engineered to exhibit altered susceptibility to a specific proprotein
CC or to produce proprotein e.g. for gene therapy. Activation by a target
CC protease rather than furin allows toxicity to be more cell-specific
CC than for PE. The proproteins also provide more specific cancer
CC treatment than previous immunotoxin-based therapies.
XX
SQ Sequence 12 AA;

Query Match 97.6%; Score 41; DB 19; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFY 8
DB 4 gergffy 10

RESULT 6
AA092233 ID AAY92233 standard; Peptide; 20 AA.
XX
AC AAY92233;
XX
DE 07-AUG-2000 (first entry)
XX
DE HLA class II binding antigen epitope peptide #422.
XX
KW Human leucocyte antigen; HLA class II; antigen epitope; pharmaceutical;
KW immune response; chronic viral disease; cancer; autoimmune disease;
KW rheumatoid arthritis; multiple sclerosis; myasthenia gravis; AIDS;
KW allograft rejection; allergy; Lyme disease; hepatitis; prostate cancer;
KW glomerulonephritis; food hypersensitivity; malaria.
XX
OS unidentified.
XX

```

PN WO961916-A1.

XX 02-DEC-1999.

XX 28-MAY-1999; 99WO-US12066.

XX 29-MAY-1998; 98US-0087192.

XX (EPIM-) EPIMUNE INC.

XX Settle A, Southwood S, Sidney J;

XX WPI; 2000-097143/08.

PT New compositions containing immunogenic peptide epitopes for various
PI HLA class II DR molecules useful for inducing helper T cell response

PS Claim 1; Page 47; 60pp; English.

CC The present invention relates to a new pharmaceutical composition
CC comprising a unit dose form of a peptide, or analogue, comprising an
CC epitope selected from those represented by peptides AY9812-Y99339
CC which are derived from various antigens for various human leucocyte
CC antigen class DR molecules, representative of the world wide population.
CC The peptide/analogue binds to an HLA class II molecule at an IC-50 of
CC less than or equal to 1,000 nM. The pharmaceutical can be used to induce
CC a helper T cell response. The pharmaceutical focuses the immune response
CC towards selected determinants and could therefore be used in cases of
CC chronic viral diseases and cancer. Examples of diseases that can be
CC treated using the peptide containing pharmaceutical include autoimmune
CC diseases (rheumatoid arthritis, multiple sclerosis, and myasthenia
CC gravis), allograft rejection, allergies, Lyme disease, hepatitis,
CC post-streptococcal endocarditis or glomerulonephritis and food
CC hypersensitivities. The peptide epitopes can be used to enhance immune
CC responses against other immunogens administered with the peptides.
CC Diseases which can be treated using immunogenic mixtures include
CC prostate cancer, hepatitis B, hepatitis C, AIDS, renal carcinoma,
CC cervical carcinoma, lymphoma, and condyloma acuminatum. The peptides may
CC also be used to make monoclonal antibodies useful as potential diagnostic
CC or therapeutic agents. The peptides may also be useful as diagnostic
CC reagents, for example, to determine the susceptibility of an individual
CC to a treatment regimen. Also, the peptides may be used to predict which
CC individuals will be at substantial risk of developing chronic infection.
CC The selection of appropriate T and B cell epitopes should allow the
CC development of epitope based vaccines particularly towards conserved
CC epitopes of pathogens which are characterized by high sequence
CC variability such as HIV, HCV and Malaria.

XX Sequence 20 AA:

Query Match 97.6%; Score 41; DB 21; Length 20;

Best Local Similarity 100.0%; Pred. No. 0.25; Mismatches 0; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFEY 8
DB 10 gergfyy 16

RESULT 7

AA048863 ID AAB48863 standard; peptide; 26 AA.

XX AAB48863;

XX 16-MAR-2001 (first entry)

DE Mutant human insulin (IA protein) cys-4 B chain.

XX Human insulin; insulin activity protein; IA protein;

KW diabetes mellitus; type 1; type 2; stability;
KW disulphide bond; cysteine replacement; mutant; mutein.

XX Homo sapiens.
OS Synthetic.

XX WO200069901-A2.

XX 23-NOV-2000.

XX 19-MAY-2000; 2000MO-US13764.

XX 19-MAY-1999; 99US-0134930.

XX (XENC-) XENCOR INC.

XX Dahlyat BI;

XX WPI; 2001-025004/03.

PT Non-naturally occurring protein with insulin activity useful for
PI treating type 1 and type 2 diabetes, comprising amino acid
PI substitutions as compared to native human insulin and having enhanced
PI stability.

PS Claim 10; Fig 3G; 95pp; English.

CC The invention relates to novel non-naturally occurring mature human
CC insulin mutants, designated insulin activity (IA) proteins in the
CC specification (AAB48850-B48876), which have altered properties (e.g.,
CC enhanced oxidative, alkaline or thermal stability, prolonged shelf-life,
CC altered ability to oligomerise) compared to wild-type mature human
CC insulin (AAB48848, AAB48849) but which are still capable of binding to
CC an insulin receptor. The insulin mutants of the invention have less than
CC 98% sequence identity to wild-type human insulin, but have a 3D
CC structure which substantially corresponds to that of wild-type human
CC insulin. The invention also relates to recombinant nucleic acids encoding
CC the insulin mutants, expression vectors and host cells comprising mutant
CC insulin-encoding DNA, the recombinant preparation of an insulin mutant,
CC and a pharmaceutical composition comprising an insulin mutant of the
CC invention. The mature human insulin mutants are useful for treating
CC insulin-responsive conditions and disorders of carbohydrate metabolism
CC such as type 1 or type 2 diabetes. The nucleic acids encoding the
CC insulin mutants are used in gene therapy techniques for treating these
CC conditions. The insulin mutants are more stable compared to wild-type
CC mature human insulin, or have an improved ability to form insulin
CC hexamers. Sequences AAB48850-B48863 represent the A and B chains of
CC insulin mutants in which one or more cysteine residue has been
CC replaced, thus preventing the formation of at least one disulphide bond
CC and improving stability and activity.

XX Sequence 26 AA:

Query Match 97.6%; Score 41; DB 22; Length 26;

Best Local Similarity 100.0%; Pred. No. 0.32; Mismatches 0; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFEY 8
DB 16 gergfyy 22

RESULT 8

AA029997 ID AAR29997 standard; Protein; 27 AA.

XX AAR29997;

XX 05-MAY-1993 (first entry)

DE Human insulin B-chain analogue (generic).

XX Insulin; diabetes; chain; activity; dimerisation; self-association.

OS	Homo sapiens.
XX	
EH	Key
FT	Misc-difference 1
FT	/label- PHE, ASP
FT	Misc-difference 3
FT	/label- ASN, ASP
FT	Disulfide-bond 7
FT	/note- "Cys7 of the A-chain is linked to
FT	Cys7 of the B-chain"
FT	Disulfide-bond 19
FT	/note- "Cys20 of the A-chain is linked to
FT	Cys19 of the B-chain"
XX	
PN	EP519750-A.
XX	
PD	23-DEC-1992.
XX	
PF	19-JUN-1992; 92EP-0305678.
XX	
PR	21-JUN-1991; 91US-0718574.
XX	
PA	(ELIL) LILLY & CO ELI.
XX	
PI	Brems DN, Chance RE, Frank BH, Chence RE;
XX	
DR	WPI; 1992-425940/52.
XX	
PT	New human insulin analogues - are hypoglycaemic agents for
XX	treatment of diabetes etc.
XX	
PS	Claim 1; Page 17; 19pp; English.
XX	
CC	The human Insulin A-chain analogue (AAR29996) is properly
CC	cross-linked (i.e. 3 disulfide bridges) to the human insulin
CC	B-chain analogue (AAR29997) or a fragment of the B-chain analogue
CC	comprising residues 1-26, 2-27 or 2-26. The wild-type His10 of
CC	the B-chain has been replaced by Asp10.
CC	The Insulin analogue is more stable and less prone to dimerisation
CC	or self-association to higher mol. wt. forms and thereby possesses
CC	a comparatively more rapid onset of activity while retaining the
XX	biological activity of native human Insulin.
XX	
Sequence	27 AA;

```

Query Match          97.6%; Score 41; DB 13; Length 27;
Best Local Similarity 100.0%; Pred. No. 0.33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0
Oy      2 GERGFEY 8
        |||||
Db      20 gergfly 26

RESULT 9
AAR30006
ID AAR30006 standard; Protein; 27 AA.
XX
AC AAR30006;
XX
DT 05-MAY-1993 (first entry)
XX
DE Human Insulin B-chain analogue (Phe(B1),Asn(B3)).
XX
KW Insulin; diabetes; chain; activity; dimerisation; self-association.
XX
OS Homo sapiens.
XX
Key Location/Qualifiers
FT Disulfide-bond 7 /note= "Cys7 of the A-chain is linked to
FT Cys7 of the B-chain"
FT

```

FT	Disulfide-bond	19	/note- "Cys20 of the A-chain is linked to
FT			Cys19 of the B-chain"
XX			
XX	EP519750-A.		
XX			
XX	23-DEC-1992.		
XX			
XX	19-JUN-1992;	92EP-0305678.	
XX			
XX	21-JUN-1991;	91US-0718574.	
XX			
PA	(ELIL) LILLY & CO ELI.		
XX			
PI	Brems DN, Chance RE, Frank BH, Chence RE;		
XX			
DR	WPI: 1992-425940/52.		
XX			
PT	New human insulin analogues - are hypoglycaemic agents for		
PT	treatment of diabetes etc.		
XX			
XX	Claim 3-4; Page 17-18; 19pp; English.		
CC			
CC	The pref. human insulin A-chain analogue (Asn21) is properly		
CC	cross-linked (i.e. 3 disulfide bridges) to the pref. human insulin		
CC	B-chain analogue (Phe1, Asn2) or a fragment of the B-chain analogue		
CC	comprising residues 1-26, 2-27 or 2-28. The wild-type His10 of		
CC	the B-chain has been replaced by Asp10.		
CC	The insulin analogue is more stable and less prone to dimerisation		
CC	or self-association to higher mol. wt. forms and thereby possesses		
CC	a comparatively more rapid onset of activity while retaining the		
XX	biological activity of native human insulin.		
XX			
Sequence	27 AA;		
50			

```

Query Match Similarity      97.6%; Score 41; DB 13; Length 27;
Best Local Similarity      100.0%; Pred. No. 0.33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

OY      2 GERGFEY 8
        |||||
Db      20 gergfy 26

RESULT 10
AAR30007
ID      AAR30007 standard; Protein; 27 AA.
XX
XX      AAR30007;
AC
XX
DT      05-MAY-1993 (first entry)
XX
DE      Human Insulin B-chain analogue (Phe(B1),Asp(B3)).
XX
XX      Insulin; diabetes; chain; activity; dimerisation; self-association.
KM
XX
OS      Homo sapiens.
XX
XX
FH      Key Location/Qualifiers
FT      Disulfide-bond 7 /note= "Cys7 of the A-chain is linked to
FT      Cys7 of the B-chain"
FT      Disulfide-bond 19 /note= "Cys20 of the A-chain is linked to
FT      Cys19 of the B-chain"
XX
XX      EP519750-A.
XX
XX      23-DEC-1992.
XX
XX      19-JUN-1992; 92EP-0305678.
XX

```

```

PR 21-JUN-1991; 91US-0718574.
XX (ELIL ) LILLY & CO ELI.
XX Brems DN, Chance RE, Frank BH, Chence RE;
XX WPI; 1992-425940/52.
XX DR
XX PT New human insulin analogues - are hypoglycaemic agents for
XX treatment of diabetes etc.
XX PS Claim 1; Page 17; 19pp; English.
XX
XX The human insulin A-chain analogue (AAR29996) is properly
XX cross-linked (i.e. 3 disulfide bridges) to the human insulin
XX B-chain analogue (Phel, Asp3) or a fragment of the B-chain analogue
XX comprising residues 1-26, 2-27 or 2-26. The wild-type His10 of
XX the B-chain has been replaced by Asp10.
XX CC The insulin analogue is more stable and less prone to dimerisation
XX or self-association to higher mol. wt. forms and thereby possesses
XX a comparatively more rapid onset of activity while retaining the
XX biological activity of native human insulin.
XX
SQ Sequence 27 AA;

Query Match 97.6%; Score 41; DB 13; Length 27;
Best Local Similarity 100.0%; Pred. No. 0.33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFY 8
DB 20 gergrfy 26

RESULT 11
AAR30008
ID AAR30008 standard; Protein; 27 AA.
XX
XX AAR30008;
XX
XX 05-MAY-1993 (first entry)
XX
XX Human insulin B-chain analogue (Asp(B1),Asn(B3)).
XX
XX Insulin; diabetes; chain; activity; dimerisation; self-association.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Disulfide-bond 7 /note= "Cys7 of the A-chain is linked to
XX Cys7 of the B-chain"
XX Disulfide-bond 19 /note= "Cys20 of the A-chain is linked to
XX Cys19 of the B-chain"
XX
XX EP519750-A.
XX 23-DEC-1992.
XX
XX 19-JUN-1992; 92EP-0305678.
XX
XX 21-JUN-1991; 91US-0718574.
XX
XX (ELIL ) LILLY & CO ELI.
XX Brems DN, Chance RE, Frank BH, Chence RE;
XX WPI; 1992-425940/52.
XX
XX New human insulin analogues - are hypoglycaemic agents for
XX treatment of diabetes etc.

```

```

XX PS Claim 1; Page 17; 19pp; English.
XX
XX The human insulin A-chain analogue (AAR29996) is properly
XX cross-linked (i.e. 3 disulfide bridges) to the human insulin
XX B-chain analogue (Asp1,Asn3) or a fragment of the B-chain analogue
XX comprising residues 1-26, 2-27 or 2-26. The wild-type His10 of
XX the B-chain has been replaced by Asp10.
XX CC The insulin analogue is more stable and less prone to dimerisation
XX or self-association to higher mol. wt. forms and thereby possesses
XX a comparatively more rapid onset of activity while retaining the
XX biological activity of native human insulin.
XX
SQ Sequence 27 AA;

Query Match 97.6%; Score 41; DB 13; Length 27;
Best Local Similarity 100.0%; Pred. No. 0.33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GERGFY 8
DB 20 gergrfy 26

RESULT 12
AAR30009
ID AAR30009 standard; Protein; 27 AA.
XX
XX AAR30009;
XX
XX 05-MAY-1993 (first entry)
XX
XX Human insulin B-chain analogue (Asp(B1),Asp(B3)).
XX
XX Insulin; diabetes; chain; activity; dimerisation; self-association.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Disulfide-bond 7 /note= "Cys7 of the A-chain is linked to
XX Cys7 of the B-chain"
XX Disulfide-bond 19 /note= "Cys20 of the A-chain is linked to
XX Cys19 of the B-chain"
XX
XX EP519750-A.
XX 23-DEC-1992.
XX
XX 19-JUN-1992; 92EP-0305678.
XX
XX 21-JUN-1991; 91US-0718574.
XX
XX (ELIL ) LILLY & CO ELI.
XX Brems DN, Chance RE, Frank BH, Chence RE;
XX WPI; 1992-425940/52.
XX
XX New human insulin analogues - are hypoglycaemic agents for
XX treatment of diabetes etc.
XX
XX Claim 1; Page 17; 19pp; English.
XX
XX The human insulin A-chain analogue (AAR29996) is properly
XX cross-linked (i.e. 3 disulfide bridges) to the human insulin
XX B-chain analogue (Asp1, Asp3) or a fragment of the B-chain analogue
XX comprising residues 1-26, 2-27 or 2-26. The wild-type His10 of
XX the B-chain has been replaced by Asp10.
XX CC The insulin analogue is more stable and less prone to dimerisation
XX or self-association to higher mol. wt. forms and thereby possesses

```

CC a comparatively more rapid onset of activity while retaining the
CC biological activity of native human insulin.
XX
SQ Sequence 27 AA;

Query Match 97.6%; Score 41; DB 13; Length 27;
Best Local Similarity 100.0%; Pred. No. 0.33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 2 GERGFY 8
| | | | | | |
Db 20 gergrfy 26

RESULT 13
AAW46931
ID AAW46931 standard; peptide: 28 AA.
XX
AC AAW46931;
XX
DT 26-JUN-1998 (first entry)
XX
DE Human Insulin B chain analogue.
XX
KM A chain; Insulin analogue; production; B chain; human;
XX proinsulin-like precursor; treatment; hyperglycaemia.
OS Synthetic.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Cross-links 5 /note= "forms a disulphide bond with the
FT A chain"
FT Cross-links 17 /note= "forms a disulphide bond with the
FT A chain"
FT Misc-difference 26 /label= P26K
FT /note= "Pro28 replaced with Lys"
FT Misc-difference 27 /label= K27P
FT /note= "Lys29 replaced with Pro"
XX
PN US5700662-A.
XX
PD 23-DEC-1997.
XX
PF 06-JUN-1995; 95US-0308352.
XX
PR 17-APR-1991; 91US-0686632.
XX
PA (ELIL) LILLY & CO ELI.
PI Chance RE, DIMarchi RD, Frank BH, Shields JE;
XX
DR WPI: 1998-062376/06.
XX
PT Production of Insulin analogues - by recombinant DNA techniques, for
XX treating hyperglycaemia
PS
XX Example 27; Page -: 49pp; English.
CC AAW46928-38 represent Insulin B chain analogues. The specification
CC describes a novel process for the production of Insulin analogues.
CC This process comprises transforming a host cell with DNA encoding a
CC proinsulin-like precursor having the A and B chains, culturing the
CC cell under conditions suitable for expressing the DNA, isolating the
CC proinsulin-like precursor, and enzymatically cleaving the precursor to
CC remove the connecting peptide. Alternatively, a host cell can be
CC transformed with DNA that individually encodes the A and B chains, which
CC are then expressed, isolated and combined. Some of the analogues

CC produced by the method have higher activity than native human Insulin.
CC The Insulin analogues produced by the method can be used for treating
CC hyperglycaemia. note: this sequence does not appear in the specification;
CC it was created using information provided.
XX
SQ Sequence 28 AA;

Query Match 97.6%; Score 41; DB 19; Length 28;
Best Local Similarity 100.0%; Pred. No. 0.35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 2 GERGFY 8
| | | | | | |
Db 18 gergrfy 24

RESULT 14
AAW46932
ID AAW46932 standard; peptide: 28 AA.
XX
AC AAW46932;
XX
DT 26-JUN-1998 (first entry)
XX
DE Human Insulin B chain analogue.
XX
KM A chain; Insulin analogue; production; B chain; human;
XX proinsulin-like precursor; treatment; hyperglycaemia.
OS Synthetic.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Cross-links 5 /note= "forms a disulphide bond with the
FT A chain"
FT Misc-difference 8 /label= H8D
FT /note= "His10 replaced with Asp"
FT Cross-links 17 /note= "forms a disulphide bond with the
FT A chain"
FT Misc-difference 26 /label= P26K
FT /note= "Pro28 replaced with Lys"
FT Misc-difference 27 /label= K27P
FT /note= "Lys29 replaced with Pro"
XX
PN US5700662-A.
XX
PD 23-DEC-1997.
XX
PF 06-JUN-1995; 95US-0308352.
XX
PR 17-APR-1991; 91US-0686632.
XX
PA (ELIL) LILLY & CO ELI.
PI Chance RE, DIMarchi RD, Frank BH, Shields JE;
XX
DR WPI: 1998-062376/06.
XX
PT Production of Insulin analogues - by recombinant DNA techniques, for
XX treating hyperglycaemia
PS
XX Example 27; Page -: 49pp; English.
CC AAW46928-38 represent Insulin B chain analogues. The specification
CC describes a novel process for the production of Insulin analogues.
CC This process comprises transforming a host cell with DNA encoding a
CC proinsulin-like precursor having the A and B chains, culturing the

CC cell under conditions suitable for expressing the DNA, isolating the
 CC proinsulin-like precursor, and enzymatically cleaving the precursor to
 CC remove the connecting peptide. Alternatively, a host cell can be
 CC transformed with DNA that individually encodes the A and B chains, which
 CC are then expressed, isolated and combined. Some of the analogues
 CC produced by the method have higher activity than native human insulin.
 CC The insulin analogues produced by the method can be used for treating
 CC hyperglycaemia. note: this sequence does not appear in the specification;
 CC it was created using information provided.

Sequence 28 AA;

Query Match 97.6%; Score 41; DB 19; Length 28;
 Best Local Similarity 100.0%; Pred. No. 0.35;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 GERGFEY 8
 Db 18 gergffy 24

RESULT 15.

AAW46924 standard; peptide; 28 AA.

AAW46924;

26-JUN-1998 (first entry)

Insulin B chain analogue.

A chain; insulin analogue; production; B chain; human;
 proinsulin-like precursor; treatment; hyperglycaemia.

Synthetic.

Location/Qualifiers
 7
 /note- "forms a disulphide bond with the
 A chain"

19
 /note- "forms a disulphide bond with the
 A chain"

US5700662-A.

23-DEC-1997.

06-JUN-1995; 95US-0308352.

17-APR-1991; 91US-0686632.

(ELIL) LILLY & CO ELI.

Chance RE, Dimarchi RD, Frank BH, Shields JE;

WPI. 1998-062376/06.

Production of insulin analogues - by recombinant DNA techniques, for
 treating hyperglycaemia

Claim 6; Page -: 49pp; English.

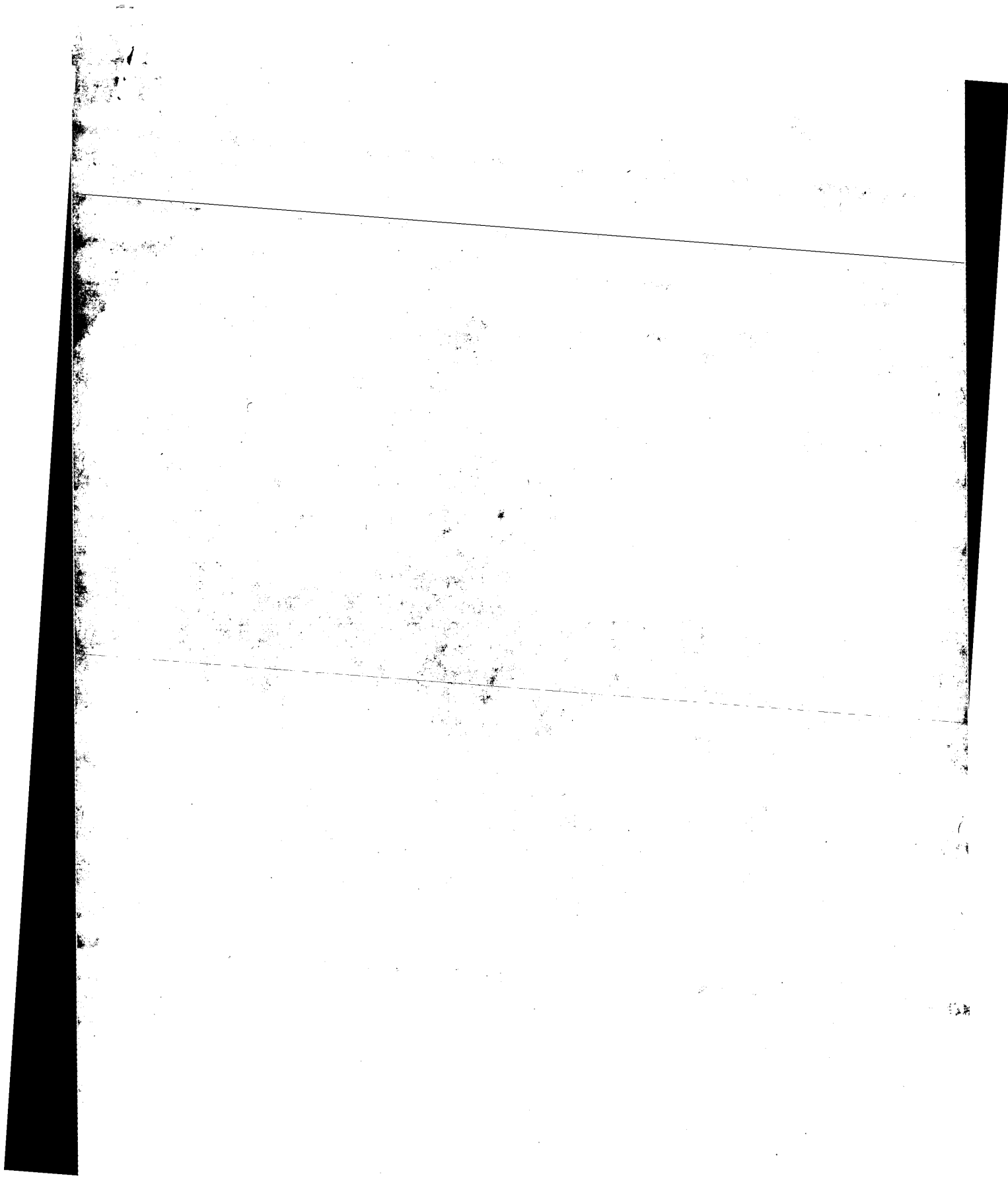
The present sequence represents an insulin B chain analogue. It is
 based on the generic sequence AAW41317, where B1 is absent, B2 is
 absent, B10 is His, B28 is Lys, B29 is Pro and B30 is Thr. The A chain
 of the insulin analogue has an Asn at position 21. The specification
 describes a novel process for the production of insulin analogues.
 This process comprises transforming a host cell with DNA encoding a
 proinsulin-like precursor having the A and B chains, culturing the
 cell under conditions suitable for expressing the DNA, isolating the
 proinsulin-like precursor, and enzymatically cleaving the precursor to

CC remove the connecting peptide. Alternatively, a host cell can be
 CC transformed with DNA that individually encodes the A and B chains, which
 CC are then expressed, isolated and combined. Some of the analogues
 CC produced by the method have higher activity than native human insulin.
 CC The insulin analogues produced by the method can be used for treating
 CC hyperglycaemia. note: this sequence does not appear in the specification;
 CC it was created using information provided.

Sequence 28 AA;

Query Match 97.6%; Score 41; DB 19; Length 28;
 Best Local Similarity 100.0%; Pred. No. 0.35;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 GERGFEY 8
 Db 18 gergffy 24

Search completed: September 6, 2001, 16:43:29
 Job time: 358 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:49:50 ; Search time 231.42 Seconds
(without alignments)
4.574 Million cell updates/sec

Title: US-09-603-713-22

Perfect score: 39
Sequence: 1 HLXSHLV 8

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP:REMBL_16:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_mhc:*
10: sp_mhc:*
11: sp_mhc:*
12: sp_mhc:*
13: sp_mhc:*
14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	92.3	108	13	Q9DDE5
2	34	87.2	223	2	P74968
3	34	87.2	417	2	Q9KYK3
4	33	84.6	418	5	Q23144
5	32	82.1	179	4	Q9NX34
6	32	82.1	368	2	Q9PCP6
7	32	82.1	382	5	Q9J3E1
8	32	82.1	833	3	Q13620
9	32	82.1	969	3	Q9V9Q2
10	32	82.1	2532	3	Q9Y7D5
11	31	79.5	310	10	Q9M823
12	31	79.5	332	10	Q9M849
13	31	79.5	431	4	Q9Z967
14	31	79.5	466	2	Q83526
15	31	79.5	579	4	Q9H5H7
16	31	79.5	582	10	Q9LGT0
17	31	79.5	1829	5	Q9VMG8
18	31	79.5	2055	4	Q75055
19	30	76.9	106	13	Q91807

20	30	76.9	171	2	Q68914	Q68914 streptomyc
21	30	76.9	263	5	Q9VE18	Q9VE18 drosophila
22	30	76.9	354	3	Q9N639	Q9N639 leishmania
23	30	76.9	356	3	Q11004	Q11004 schizosacch
24	30	76.9	375	3	Q9P3X9	Q9P3X9 neurospora
25	30	76.9	404	10	Q9X1L2	Q9X1L2 arabidopsis
26	30	76.9	431	2	Q915C7	Q915C7 pseudomonas
27	30	76.9	521	2	Q24867	Q24867 helicobacte
28	30	76.9	523	2	Q9ZK39	Q9ZK39 helicobacte
29	30	76.9	570	10	Q9LY75	Q9LY75 arabidopsis
30	30	76.9	581	14	Q84694	Q84694 pea enation
31	30	76.9	618	10	Q9M4B7	Q9M4B7 arabidopsis
32	30	76.9	618	10	Q9LKL2	Q9LKL2 arabidopsis
33	30	76.9	703	13	Q93618	Q93618 brachydanio
34	30	76.9	791	2	Q9KK17	Q9KK17 yersinia en
35	30	76.9	1173	10	Q9LFRO	Q9LFRO arabidopsis
36	30	76.9	1476	13	Q90285	Q90285 carassius a
37	30	76.9	1849	14	Q9JGK9	Q9JGK9 sagiyama vi
38	30	76.9	2467	14	Q9JGL0	Q9JGL0 sagiyama vi
39	30	76.9	5107	5	Q24498	Q24498 drosophila
40	29	74.4	89	2	Q9RC98	Q9RC98 bacillus ha
41	29	74.4	94	6	Q46425	Q46425 oryctolagus
42	29	74.4	122	4	Q9H2D7	Q9H2D7 homo sapien
43	29	74.4	229	5	Q9V808	Q9V808 drosophila
44	29	74.4	233	2	Q9FC25	Q9FC25 streptomyce
45	29	74.4	236	3	Q9HEC7	Q9HEC7 neurospora

ALIGNMENTS

RESULT	ID	Q9DDE5	PRELIMINARY:	PRT:	108 AA.
AC	Q9DDE5	01-MAR-2001 (TREMBLrel. 16, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)				
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)				
DE	INSULIN PRECURSOR.				
GN	INS.				
OS	Brachydanio rerio (zebrafish) (zebra danio).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;				
OC	Cypriniformes; Cyprinidae; Rasbora; Danio.				
OX	NCBI_TaxID=7955;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=9425190; PubMed=10495291;				
RA	Argenton F., Zecchin E., Bortolussi M.;				
RT	*Early appearance of pancreatic hormone-expressing cells in the				
RT	zebrafish embryo.*;				
RL	Mech. Dev. 87:217-221(1999).				
DR	EMBL, AF237750; CAC20109.1; -.				
KW	Signal.				
FT	SIGNAL	1	23	POTENTIAL.	
FT	CHAIN	24	53	INSULIN B CHAIN.	
FT	CHAIN	86	108	INSULIN A CHAIN.	
SO	SEQUENCE	108 AA;	11904 MW;	3195289E72AD6D25 CRC64;	

Query Match 92.3%; Score 36; DB 13; Length 108;
Best local Similarity 87.5%; Pred. No. 2.6;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	HLXSHLV 8
DB	28	HLXSHLV 35

RESULT	2	PRELIMINARY:	PRT:	223 AA.
ID	P74968			
AC	P74968;			

DT 01-FEB-1997 (TREMBLrel. 02, Created)
 DT 01-FEB-1997 (TREMBLrel. 02, last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)
 DE RESPONSE REGULATOR OF TWO-COMPONENT REGULATORY SYSTEM.
 GN RHPR.
 OS Wolinella succinogenes.
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 CC Wolinella.
 NCBI_TaxID=844;
 OX NCBI_TaxID=844;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Simon J., Krafft T., Klimmek O., Gross R., Kroger A.;
 RL Submitted (NOV-1996) to the EMBL/Genbank/DBJ databases.
 CC -1- SIMILARITY: THE N-TERMINAL REGION IS SIMILAR TO THAT OF OTHER
 CC REGULATORY COMPONENTS OF SENSOR TRANSDUCTION SYSTEMS.
 CC -1- SIMILARITY: TO BACTERIAL REGULATORY PROTEINS INVOLVED IN SIGNAL
 CC TRANSDUCTION.
 DR EMBL; Y09296; CAA70482.1; -
 DR HSSP; P41789; INTR.
 DR InterPro: IPR001789; -
 DR InterPro: IPR001667; -
 DR Pfam; PF00072; response_reg.1.
 DR Pfam; PF00486; trans_reg_C.1.
 DR SMART; SM00448; REC.1.
 KW DNA-binding; Phosphorylation; Sensory transduction;
 KW Transcription regulation;
 SO SEQUENCE 223 AA; 25562 MW; EB02D688C1F383A6 CRC64;

Query Match 87.2%; Score 34; DB 2; Length 223;
 Best Local Similarity 75.0%; Pred. No. 14;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 HLXGSHLV 8
 11111111
 DB 39 HLDGSHLL 46

RESULT 3
 Q9KYK3 PRELIMINARY; PRT; 417 AA.
 AC Q9KYK3;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)
 DE PROBABLE ACYL-COA DEHYDROGENASE.
 GN SC168.10C.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
 NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Saunders D.C., Harris D.;
 RC STRAIN-A3(2);
 RL Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA STRAIN-A3(2);
 RC Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA STRAIN-A3(2);
 RC MEDLINE-97000351; PubMed-8843436;
 RA Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RA "A set of ordered cosmids and a detailed genetic and physical map for
 RL the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 DR MOL. Microbiol. 21:77-96(1996).
 DR EMBL; AL356593; CAB92227.1; -
 DR InterPro: IPR001552; -
 DR Pfam; PF00441; Acyl-CoA_dh; 1.

SQ SEQUENCE 417 AA; 44474 MW; 8092222E685DF640 CRC64;

Query Match 87.2%; Score 34; DB 2; Length 417;
 Best Local Similarity 62.5%; Pred. No. 27;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 HLXGSHLV 8
 11111111
 DB 172 HLAGAHLI 179

RESULT 4
 Q23144 PRELIMINARY; PRT; 418 AA.
 AC Q23144;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, last annotation update)
 DE W03C9.2 PROTEIN.
 GN W03C9.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Gregory J., Ainscough R.;
 RL Submitted (OCT-1995) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copey J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 368:32-38(1994).
 DR EMBL; Z66516; CAA91360.1; -
 SO SEQUENCE 418 AA; 47868 MW; 54BE848DE8A51B6C CRC64;

Query Match 84.6%; Score 33; DB 5; Length 418;
 Best Local Similarity 62.5%; Pred. No. 44;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 HLXGSHLV 8
 11111111
 DB 178 HLDGNHLI 185

RESULT 5
 Q9NX34 PRELIMINARY; PRT; 179 AA.
 AC Q9NX34;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, last annotation update)
 DE CDNA FLJ20465 F15; CLONE KAT06236.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,

RA Suzuki Y., Obayashi M., Nishi T., Shibahara T., Tanaka T.,
 RA Nakamura Y., Isogai T., Sugano S.,
 RT "NEBO human cDNA sequencing project";
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK000472; BAA91188.1; -;
 SO SEQUENCE 179 AA; 19372 MW; DSCDA2F8F2368585 CRC64;

Query Match 82.1%; Score 32; DB 4; Length 179;
 Best Local Similarity 62.5%; Pred. No. 30;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 HLXGSHLV 8
 DB 30 HISGCHLV 37

RESULT 6
 O9PCP6 PRELIMINARY; PRT; 368 AA.
 AC O9PCP6;
 DT 01-OCT-2000 (TReMBLrel. 15, Created)
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
 DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
 DE NAD(P)H-DEPENDENT 2-CYCLOHEXEN-1-ONE REDUCTASE.
 GN Xyl1732.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 OC Xylella.
 NC NCBITaxid=23711;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9A5C;
 RX MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G., Relnack F.C., Arruda P., Abreu F.A., Acencio N.,
 RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carier H.,
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Doroty H.,
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,
 RA Garner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hohnselt J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Martino C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Montelero-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Peloto B.R., Pereira G.A.G., Pereira H.A.Jr., Pesquero J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E.Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A.Jr.,
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Trufi D., Tsai S.M., Tsubaki M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
 RT "The genome sequence of the plant pathogen xylella fastidiosa";
 RL Nature 406:151-159(2000).
 DR EMBL: AE003996; AAF84541.1; -;
 DR InterPro: IPR001155; -;
 DR Pfam: PF00724; oxidored_FMN.1.
 SO SEQUENCE 368 AA; 39094 MW; C3B4668888A7D8D CRC64;

Query Match 82.1%; Score 32; DB 2; Length 368;
 Best Local Similarity 62.5%; Pred. No. 62;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 HLXGSHLV 8

DB 98 HLXGSHLV 105

RESULT 7
 O9U3E1 PRELIMINARY; PRT; 382 AA.
 AC O9U3E1;
 DT 01-MAY-2000 (TReMBLrel. 13, Created)
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
 DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
 DE F58A4.7B PROTEIN.
 GN F58A4.7B.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 NC NCBITaxid=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Berts M.;
 RL Submitted (Apr-1993) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RT investigating biology";
 RL Science 282:2012-2018(1998).
 DR EMBL: Z22179; CAA80170.1; -;
 DR InterPro: IPR001092; -;
 DR InterPro: IPR003015; -;
 DR Pfam: PF00010; HLH; 1.
 DR PROSITE: PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
 DR SMART: SM00353; HLH; 1.
 SO SEQUENCE 382 AA; 40763 MW; 510BEF225B073804 CRC64;

Query Match 82.1%; Score 32; DB 5; Length 382;
 Best Local Similarity 62.5%; Pred. No. 65;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 HLXGSHLV 8
 DB 361 HLEGSFHI 368

RESULT 8
 O13620 PRELIMINARY; PRT; 833 AA.
 AC O13620;
 DT 01-JAN-1998 (TReMBLrel. 05, Created)
 DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
 DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
 DE HYPOTHETICAL 93.7 KDA PROTEIN.
 GN P1029.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 NC NCBITaxid=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972 H-;
 RA Kushida N., Yamazaki S., Tanaka T., Jinno K., Halkawa Y., Yamazaki J.,
 RA Yamamoto S., Sekine M., Oguchi A., Nagai Y., Sakai M., Aoki K.,
 RA Ogura K., Otsuka R., Kudoh Y., Yanagida M., Machida M., Zhang M.Q.;
 RL Submitted (May-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB004535; BAA21408.1; -;
 DR HSSP: P09012; 1FHT.
 DR InterPro: IPR000504; -;
 DR Pfam: PF00076; trm; 5.
 DR PROSITE: PS00030; RNP_1; UNKNOWN_2.
 DR SMART: SM00360; RRM; 1.

KW Hypothetical protein.
SQ SEQUENCE 833 AA; 93676 MW; 64FE767D43E02FE4 CRC64;

Query Match
Best Local Similarity 82.1%; Score 32; DB 3; Length 833;
Matches 6; Conservative 0; Pred. No. 1.5e+02; Mismatches 2; Indels 0; Gaps 0;

OY 1 HLKSHLV 8
DB 785 HLKSHLV 792

RESULT 9
ID O9V902 PRELIMINARY; PRT: 969 AA.
AC O9V902: 01-MAY-2000 (TREMBlrel. 13; Created)
DT 01-MAY-2000 (TREMBlrel. 13; Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16; Last annotation update)
DE CG11631 PROTEIN.
GN CG11631.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Eukaryota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydriidae; Drosophilidae; Drosophila.
OC NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MELLINE-20196006; Pubmed=10731132;
RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amaratilake P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang O., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abell J.F., Agbayani A., An H.-J., Andrews-Plannoch C., Baldwin D.,
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Delhke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischman W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Hostlin D., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Jaitli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Jaitli B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasro P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paolel J.M.,
RA Palazolo M., Pittman G.S., Saunders R.D.C., Scheeler F., Shen H.,
RA Reinert K., Remington K., Simpson M., Skungski M.P., Smith T.,
RA Shue B.C., Siden-Kiamos I., Stapleton M., Strung R., Sun E.,
RA Svirskas R., Tector C., Turner R., Ventler E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL: A003781; AAF57234.1;
DR HSSP: P08046.1A11
DR FLYBase: FBgn0032963; CG11631.
DR InterPro: IPR000822;

DR Pfam: PF00096; zf-C2H2; 8.
DR PROSITE: PS00028; ZINC_FINGER_C2H2; 4.
DR SMART: SM00355; ZNF_C2H2; 1.
DR DNA-binding: Metal-binding; Zinc-finger.
SQ SEQUENCE 969 AA; 110840 MW; 2A31C500FD497FAC CRC64;

Query Match
Best Local Similarity 82.1%; Score 32; DB 5; Length 969;
Matches 5; Conservative 1; Pred. No. 1.7e+02; Mismatches 2; Indels 0; Gaps 0;

OY 1 HLKSHLV 8
DB 367 HLKSHLV 374

RESULT 10
ID O9Y7D5 PRELIMINARY; PRT: 2532 AA.
AC O9Y7D5: 01-NOV-1999 (TREMBlrel. 12; Created)
DT 01-NOV-1999 (TREMBlrel. 12; Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16; Last annotation update)
DE POLYKETIDE SYNTHASE.
GN LOVE.
OS Aspergillus terreus.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eukaryota; Fungi; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OC NCBI_TaxID=33178;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC20542;
RA Kennedy J., Anclair K., Kendrew S.G., Park C., Vederas J.C.,
RA Hutchison C.R.;
RT "Accessory Proteins Modulate Polyketide Synthase Activity During Lovastatin Biosynthesis";
RL Science 0:0-0(1999).
DR EMBL: AF141925; AAD34559.1;
DR InterPro: IPR000051;
DR InterPro: IPR000255;
DR InterPro: IPR000794;
DR InterPro: IPR001227;
DR InterPro: IPR001601;
DR InterPro: IPR002085;
DR Pfam: PF00107; adh_zinc; 1.
DR Pfam: PF00109; ketoacyl-synt; 1.
DR Pfam: PF00550; pp-binding; 1.
DR Pfam: PF00698; Acyl_transf; 1.
DR PROSITE: PS00075; ACP_DOMAIN; 1.
DR PROSITE: PS00606; B_KETOACYL_SYNTHASE; 1.
DR PROSITE: PS00012; PHOSPHOPANTHETINE; UNKNOWN; 1.
DR PROSITE: PS00012; PHOSPHOPANTHETINE; transferase.
KW Phosphopantetheine; transferase.
SQ SEQUENCE 2532 AA; 276638 MW; C486622B89D58B2E CRC64;

Query Match
Best Local Similarity 82.1%; Score 32; DB 3; Length 2532;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 HLKSHLV 8
DB 973 HVGSHLV 980

RESULT 11
ID O9M8Z3 PRELIMINARY; PRT: 310 AA.
AC O9M8Z3: 01-OCT-2000 (TREMBlrel. 15; Created)
DT 01-OCT-2000 (TREMBlrel. 15; Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15; Last annotation update)
DE T6K12.1 PROTEIN (FRAGMENT).
GN T6K12.1.

OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B., Wu D.,
 RA Rongling C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
 RA Bowman C.L., White O., Niernan W.C., Fraser C.M.;
 RT Arabidopsis thaliana chromosome III BAC T6K12 genomic sequence.;
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC016829; AAF26777.1;
 FT NON_TER 310 310
 SO SEQUENCE 310 AA; 34219 MW; 71B0DCA6D3A8D5CD CRC64;

Query Match 79.5%; Score 31; DB 10; Length 310;
 Best Local Similarity 62.5%; Pred. No. 84;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 HLXGSHLV 8
 DB 136 HLKCHLI 143

RESULT 12
 O9M849 PRELIMINARY; PRT; 332 AA.
 AC O9M849;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE T27C4.1 PROTEIN.
 GN T27C4.1.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B., Wu D.,
 RA Rongling C.M., Koo H., Fujii C.Y., Utterback T.R.,
 RA Barnstead M.E., Bowman C.L., White O., Niernan W.C., Fraser C.M.;
 RT Arabidopsis thaliana chromosome III BAC T27C4 genomic sequence.;
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC022287; AAF63768.1;
 SO SEQUENCE 332 AA; 36788 MW; 69CA656186BFAE5B CRC64;

Query Match 79.5%; Score 31; DB 10; Length 332;
 Best Local Similarity 62.5%; Pred. No. 91;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 HLXGSHLV 8
 DB 136 HLKCHLI 143

RESULT 13
 O92967 PRELIMINARY; PRT; 431 AA.
 AC O92967;
 DT 01-FEB-1997 (TREMBlrel. 02, Created)
 DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE ZINC FINGER PROTEIN ZFP6 (FRAGMENT).
 GN ZF6.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Petroni D., Bartolini E., Ottolenghi S., Comi P.;
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U71363; AAB16809.1;
 DR HSSP: P25490; 12NM.
 DR InterPro: IPR000822;
 DR Pfam: PF00096; zf-C2H2; 9.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2; 9.
 DR SMART: SM00353; Znf_C2H2; 1.
 KW DNA-binding; Metal-binding; Zinc-finger.
 FT NON_TER 431 431
 SO SEQUENCE 431 AA; 49121 MW; CBCE47D1EB530F07 CRC64;

Query Match 79.5%; Score 31; DB 4; Length 431;
 Best Local Similarity 71.4%; Pred. No. 1,2e+02;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 HLXGSHLV 7
 DB 114 HVGSHLV 120

RESULT 14
 O83526 PRELIMINARY; PRT; 466 AA.
 AC O83526;
 DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE K+ TRANSPORT PROTEIN (TRKA).
 GN TP0513.
 OS Treponema pallidum.
 OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
 OC NCBI_TaxID=160;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NICHOLS;
 RX MEDLINE=98332770; PubMed=9665876;
 RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
 RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
 RA Sledge R., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
 RA Kralak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
 RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
 RA Venter J.C.;
 RT "Complete genome sequence of Treponema pallidum, the syphilis
 agent spirochete";
 RL Science 281:375-388(1998).
 DR EMBL: AE001227; AAC65501.1;
 DR TIGR: TP0513;
 DR InterPro: IPR000309;
 DR InterPro: IPR001148;
 DR Pfam: PF02080; TRKA; 2.
 DR Pfam: PF02254; KTN; 2.
 SO SEQUENCE 466 AA; 49641 MW; 7D266ED62B7F49D0 CRC64;

Query Match 79.5%; Score 31; DB 2; Length 466;
 Best Local Similarity 71.4%; Pred. No. 1,3e+02;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 HLXGSHLV 7
 DB 376 HLAGSHV 382

RESULT 15
 O9H5H7 PRELIMINARY; PRT; 579 AA.
 ID O9H5H7

AC 09H5H7: 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, last annotation update)
 DE CDNA: FLJ23425 FIS, CLONE HEP22862.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kawabata A., Hiki J., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
 RA Okitani R., Ota T., Suzuki Y., Odayashi M., Nishi T., Shibahara T.,
 RA Tanaka T., Nakamura Y., Isogai T., Sugano S.:
 RT "NEO human cDNA sequencing project."
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK027078; BAB15649.1; -
 SQ SEQUENCE 579 AA; 64639 MW; 1A2C1570D42CA17F CRC64;

Query Match 79.5%; Score 31; DB 4; Length 579;
 Best Local Similarity 62.5%; Pred. No. 1.6e+02;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 HLXGSHLV 8
 1:11111
 DB 204 HVLGAHLV 211

Search completed: September 6, 2001, 16:49:51
 Job time: 735 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:51:08 ; Search time 72.75 Seconds
(without alignments)
3.767 Million cell updates/sec

Title: US-09-603-713-22

Perfect score: 39
Sequence: 1 HLXGSHLV 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues
Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	92.3	50	1	INS2_BATSP
2	36	92.3	50	1	INS_KATPE
3	36	92.3	50	1	INS_MYOSC
4	36	92.3	50	1	INS_ONCGO
5	36	92.3	51	1	INS1_BATSP
6	36	92.3	51	1	INS2_THUTH
7	36	92.3	51	1	INS_ACOCA
8	36	92.3	51	1	INS_ANGRO
9	36	92.3	51	1	INS_ANGAN
10	36	92.3	51	1	INS_BALBO
11	36	92.3	51	1	INS_BALPH
12	36	92.3	51	1	INS_CAMDR
13	36	92.3	51	1	INS_CAPHI
14	36	92.3	51	1	INS_CHIBR
15	36	92.3	51	1	INS_DIDMA
16	36	92.3	51	1	INS_ELEMA
17	36	92.3	51	1	INS_FELCA
18	36	92.3	51	1	INS_GADCA
19	36	92.3	51	1	INS_HYSCR
20	36	92.3	51	1	INS_ORMAN
21	36	92.3	51	1	INS_TRASC
22	36	92.3	52	1	INS_ACIGU
23	36	92.3	52	1	INS_AMICA
24	36	92.3	52	1	INS_LEPSP
25	36	92.3	52	1	INS_PIAME
26	36	92.3	54	1	INS_SQUAC
27	36	92.3	57	1	INS_PETMA
28	36	92.3	70	1	INS_TORMA
29	36	92.3	81	1	INS_ANAPL
30	36	92.3	86	1	INS_HORSE
31	36	92.3	103	1	INS_SELRF
32	36	92.3	105	1	INS_BOVIN
33	36	92.3	105	1	INS_ONCKE

34	36	92.3	105	1	INS_SHEEP	P01318 ovls aries
35	36	92.3	106	1	INS1_XENLA	P12706 xenopus lae
36	36	92.3	106	1	INS2_XENLA	P12707 xenopus lae
37	36	92.3	107	1	INS_CHICK	P01332 gallus gall
38	36	92.3	108	1	INS_BRARE	O73727 brachydanto
39	36	92.3	108	1	INS_CYPCA	P01335 cyprinus ca
40	36	92.3	108	1	INS_PIG	P01315 sus scrofa
41	36	92.3	108	1	INS_RODSP	P21563 rodentia sp
42	36	92.3	110	1	INS2_MOUSE	P01326 mus musculu
43	36	92.3	110	1	INS2_RAT	P01323 rattus norv
44	36	92.3	110	1	INS_CANFA	P01321 canis fami
45	36	92.3	110	1	INS_CERAE	P30407 cercopithec

ALIGNMENTS

RESULT 1
ID INS2_BATSP STANDARD; PRT; 50 AA.
AC P01338;
DT 21-JUL-1986 (rel. 01, Created)
DT 21-JUL-1986 (rel. 01, Last sequence update)
DT 01-FEB-1995 (rel. 31, Last annotation update)
DE INSULIN 2.
OS Batrachoididae sp. (Toadfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorphia; Paracanthopterygii; Batrachoididae.
OX NCBI_TaxID=8066;
RN (1)
RP MEDLINE=66160119; PubMed=5949593;
RA Smith L.F.;
Species variation in the amino acid sequence of Insulin.;
RT Am. J. Med. 40:662-666(1966).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS, AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC PIR: A01605; INTC2.
DR HSSP: P01315; 9INS.
DR InterPro: IPR000739; .
DR Pfam: PF00049; Insulin; 1.
DR PRINTS: PR00276; INSULIN.
DR PROSITE: PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 29 INSULIN B CHAIN.
FT NON_CONS 29 30
FT CHAIN 30 50 INSULIN A CHAIN.
FT DISULFID 8 36 INTERCHAIN.
FT DISULFID 20 49 INTERCHAIN.
FT DISULFID 35 40
SQ SEQUENCE 50 AA; 5652 MW; 903E8AACBD62137C CRC64;

Query Match 92.3%; Score 36; DB 1; Length 50;
Best Local Similarity 87.5%; Pred. No. 0.41;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HLXGSHLV 8
DB 6 HLXGSHLV 13

RESULT 2
ID INS_KATPE STANDARD; PRT; 50 AA.

```

AC P01340;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE INSULIN.
GN INS.
OS Katsuwonus pelamis (Skipjack tuna) (Bonito).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Scombroidei;
OC Scombridae; Euthynnus.
OX NCBI_TaxID=8226;
RN [1]
RP SEQUENCE OF 1-29.
RA Kotaki A.;
RT "Studies on insulin. V. On the structure of the glycy chain of
RT bonito insulin II."
RL J. Biochem. 53:61-70(1963).
RN [2]
RP SEQUENCE OF 30-50.
RA Kotaki A.;
RT "Studies on insulin. III. On the structure of the alanyl chain of
RT bonito insulin."
RL J. Biochem. 51:301-309(1962).
RN [1]
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR PIR: A01607; INBN2.
DR HSSP: P01308; ILPH.
DR InterPro: IPR000739; -.
DR Pfam: PF00049; Insulin; 1.
DR PRINTS: PR00277; INSULINB.
DR PROSITE: PS00262; INSULIN. 1.
DR Kunitz family; Hormone; Glucose metabolism.
FT CHAIN 1 29 INSULIN B CHAIN.
FT NON_CONS 29 30
FT CHAIN 30 50 INSULIN A CHAIN.
FT DISULFID 7 36 INTERCHAIN.
FT DISULFID 19 49 INTERCHAIN.
FT DISULFID 35 40
SO SEQUENCE 50 AA; 5697 MW; 3627578FE24CE92E CRC64;

Query Match 92.3%; Score 36; DB 1; Length 50;
Best Local Similarity 87.5%; Pred. No. 0.41;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HLXGSHLV 8
DB 5 HLCGSHLV 12

RESULT 3
INS_MYOSC
ID INS_MYOSC STANDARD; PRT; 50 AA.
AC P07453;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE INSULIN.
GN INS.
OS Myoxocephalus scorpius (Shorthorn sculpin) (Daddy sculpin).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Scorpaeniformes;
OC Cottidae; Cottidae; Myoxocephalus.
OX NCBI_TaxID=8097;
RN [1]
RP SEQUENCE.

```

```

RP SEQUENCE.
RX MEDLINE=66274667; PubMed=3525155;
RA Cutfield J.F., Cutfield S.M., Carne A., Emidin S.O., Falkner S.;
RT "The isolation, purification and amino-acid sequence of insulin from
RT the teleost fish cottus scorpius (daddy sculpin).";
RL Eur. J. Biochem. 158:117-123(1986).
RN [1]
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR PIR: A25061; INFIS.
DR HSSP: P01308; ILPH.
DR InterPro: IPR000739; -.
DR Pfam: PF00049; Insulin; 1.
DR PRINTS: PR00276; INSULINA.
DR PRINTS: PR00277; INSULINB.
DR PROSITE: PS00262; INSULIN. 1.
DR Kunitz family; Hormone; Glucose metabolism.
FT CHAIN 1 29 INSULIN B CHAIN.
FT NON_CONS 29 30
FT CHAIN 30 50 INSULIN A CHAIN.
FT DISULFID 7 36 INTERCHAIN.
FT DISULFID 19 49 INTERCHAIN.
FT DISULFID 35 40
SO SEQUENCE 50 AA; 5682 MW; 0A600B9BEFE15827 CRC64;

Query Match 92.3%; Score 36; DB 1; Length 50;
Best Local Similarity 87.5%; Pred. No. 0.41;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HLXGSHLV 8
DB 5 HLCGSHLV 12

RESULT 4
INS_ONCGO
ID INS_ONCGO STANDARD; PRT; 50 AA.
AC P23187;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE INSULIN.
GN INS.
OS Oncorhynchus gorbuscha (Pink salmon) (Humpback salmon), and
OS Oncorhynchus kisutch (Coho salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX NCBI_TaxID=8017, 8019;
RN [1]
RP SEQUENCE.
RC SPECIES=O.gorbuscha;
RA Ruskov Y.I., Karasev V.S., Pertseva M.N., Pankov Y.A.;
RT "Amino acid sequence of humpback salmon (Oncorhynchus gorbuscha)
RT insulin."
RL Biochimica 52:247-254(1987).
RN [2]
RP SEQUENCE.
RC SPECIES=O.gorbuscha;
RX MEDLINE=90235564; PubMed=2184990;
RA Ruskov Y.I., Karasev V.S., Bondareva V.M., Pertseva M.N.,
RA Pankov Y.A.;
RT "Isolation, primary structure, and biological and immunological
RT properties of pink and chum salmon insulins."
RL Comp. Biochem. Physiol. 95B:477-482(1990).
RN [3]
RP SEQUENCE.

```

```

RC SPECIES=O.kisutch;
RX MEDLINE=85299536; Pubmed-3898237;
RA Piletskyaya E., Pollock H.G., Rouse J.B., Hamilton J.W., Kimmel J.R.,
  Gordan A.;
  Characterization of coho salmon (Oncorhynchus kisutch) insulin.;
  Regul. Pept. 11:105-116(1985).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
  INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
  FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
  CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
  DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC PIR: S02203; INON.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR PIR: A60523; INONC.
DR PIR: A60523; A60523.
DR PIR: B60523; B60523.
DR HSSP: P01308; ILPH.
DR InterPro: IPR000739; -.
DR Pfam: PF00049; Insulin; 1.
DR PRINTS: PR00276; INSULIN.
DR PROSITE: PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism.
KW CHAIN 1 29 INSULIN B CHAIN.
FT NON_CONS 29 30
FT CHAIN 30 50 INSULIN A CHAIN.
FT DISULFID 7 36 INTERCHAIN (BY SIMILARITY).
FT DISULFID 19 49 INTERCHAIN (BY SIMILARITY).
FT DISULFID 35 40 BY SIMILARITY.
SQ SEQUENCE 50 AA; 5576 MW; D3D01633158CD95F CRC64;

Query Match 92.3%; Score 36; DB 1; Length 50;
Best Local Similarity 87.5%; Pred. No. 0.41;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HLKSHLV 8
Db 5 HLKSHLV 12

RESULT 5
INS1_BATSP STANDARD; PRT; 51 AA.
AC P01337;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE INSULIN 1.
OS Batrachoidae sp. (Toadfish).
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
  Acanthomorpha; Paracanthopterygii; Batrachoidae.
OX NCBI_TaxID=8066;
RN [1]
RP SEQUENCE.
RX MEDLINE=66160119; Pubmed-5949593;
RA Smith L.F.;
  "Species variation in the amino acid sequence of insulin.";
  Am. J. Med. 40:662-666(1966).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
  INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
  FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
  CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
  DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR PIR: A01604; INTFL.
DR HSSP: P01308; ILPH.
DR InterPro: IPR000739; -.

```

```

DR Pfam: PF00049; Insulin; 1.
DR PRINTS: PR00276; INSULIN.
DR PROSITE: PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism.
KW CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 8 37 INTERCHAIN.
FT DISULFID 20 50 INTERCHAIN.
FT DISULFID 36 41
SQ SEQUENCE 51 AA; 5776 MW; A51E0BA42483705A CRC64;

Query Match 92.3%; Score 36; DB 1; Length 51;
Best Local Similarity 87.5%; Pred. No. 0.42;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HLKSHLV 8
Db 6 HLKSHLV 13

RESULT 6
INS2_THUTH STANDARD; PRT; 51 AA.
ID INS2_THUTH
AC P01339;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE INSULIN 2.
OS Thunnus thynnus (Bluefin tuna).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
  Acanthomorpha; Acanthopterygii; Perciformes; Scombroidei;
  Scombridae; Thunnus.
OX NCBI_TaxID=8237;
RN [1]
RP SEQUENCE.
RX MEDLINE=72047039; Pubmed-5406482;
RA Neumann P.A., Hummel R.E.;
  "Isolation of a single component of fish insulin from a bonito-tuna-
  swordfish insulin mixture and its complete amino-acid sequence.";
  Int. J. Protein Res. 1:125-140(1969).
RN [2]
RP IDENTIFICATION OF PROTEIN.
RA Neumann P.A., Hummel R.E.;
  Submitted (Aug-1970) to the PIR data bank.
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
  INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
  FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
  CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
  DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR PIR: A01606; INTU2.
DR HSSP: P01308; ILPH.
DR InterPro: IPR000739; -.
DR Pfam: PF00049; Insulin; 1.
DR PRINTS: PR00276; INSULIN.
DR PROSITE: PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism.
KW CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 8 37 INTERCHAIN.
FT DISULFID 20 50 INTERCHAIN.
FT DISULFID 36 41
SQ SEQUENCE 51 AA; 5742 MW; A516B9A5CBAE605A CRC64;

```

Query Match 92.3%; Score 36; DB 1; Length 51;
 Best Local Similarity 87.5%; Pred. No. 0.42;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HLXGSHLV 8
 II IIIII
 Db 6 HLCGSHLV 13

RESULT 7

INS_ACOCOA STANDARD; PRT; 51 AA.
 ID INS_ACOCOA
 AC P01324;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE INSULIN.
 GN INS.
 OS Acomys cahirinus (Egyptian spiny mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Acomys.
 OX NCBI_TaxID=10068;
 RN [1]
 RP COMPOSITION
 RX MEDLINE=72189454; PubMed=5028210;
 RA Buenzli H.F., Humbel R.E.;
 RT "Isolation and partial structural analysis of insulin from mouse (Mus musculus) and spiny mouse (Acomys cahirinus).";
 RL Hoppa-Seiler's 2. Physiol. Chem. 353:444-450(1972).
 CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
 CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO DISULFIDE BONDS.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
 DR HSP; A01591; INMSP.
 DR HSP; P01308; IBEN.
 DR InterPro: IPR000739; -.
 DR Pfam: PF00049; Insulin; 1.
 DR PRINTS: PRO0276; INSULIN.
 DR PROSITE: PS00262; INSULIN; 1.
 DR PROSITE: PS00262; INSULIN; 1.
 KW Insulin family; Hormone; Glucose metabolism.
 FT CHAIN 1 30 INSULIN B CHAIN.
 FT NON_CONS 30 31
 FT CHAIN 31 51 INSULIN A CHAIN.
 FT DISULFID 7 37 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 19 50 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 36 41
 DT SEQUENCE 51 AA; 5768 MW; 992BB8B629047D3D CRC64;

Query Match 92.3%; Score 36; DB 1; Length 51;
 Best Local Similarity 87.5%; Pred. No. 0.42;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HLXGSHLV 8
 II IIIII
 Db 5 HLCGSHLV 12

RESULT 8

INS_ANGRO STANDARD; PRT; 51 AA.
 ID INS_ANGRO
 AC P42633;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE INSULIN.
 GN INS.
 OS Anguilla rostrata (American eel).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Anguilliformes; Anguillidae;
 OC Anguillidae; Anguilla.
 OX NCBI_TaxID=7938;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Pancreas;
 RX MEDLINE=91340068; PubMed=1874385;
 RA Conlon J.M., Andrews P.C., Thim L., Moon T.W.;
 RT "The primary structure of glucagon-like peptide but not insulin has been conserved between the American eel, *Anguilla rostrata* and the European eel, *Anguilla anguilla*.";
 RL Gen. Comp. Endocrinol. 82:23-32(1991).

CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
 CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO DISULFIDE BONDS.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
 DR HSP; P01308; IHIS.
 DR InterPro: IPR000739; -.
 DR Pfam: PF00049; Insulin; 1.
 DR PRINTS: PRO0276; INSULIN.
 DR PROSITE: PS00262; INSULIN; 1.
 KW Insulin family; Hormone; Glucose metabolism.
 FT CHAIN 1 30 INSULIN B CHAIN.
 FT NON_CONS 30 31
 FT CHAIN 31 51 INSULIN A CHAIN.
 FT DISULFID 7 37 INTERCHAIN.
 FT DISULFID 19 50 INTERCHAIN.
 FT DISULFID 36 41
 DT SEQUENCE 51 AA; 5652 MW; 1999FD7BEA173CB2 CRC64;

Query Match 92.3%; Score 36; DB 1; Length 51;
 Best Local Similarity 87.5%; Pred. No. 0.42;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HLXGSHLV 8
 II IIIII
 Db 5 HLCGSHLV 12

RESULT 9

INS_ANSAN STANDARD; PRT; 51 AA.
 ID INS_ANSAN
 AC P07454; Q10995;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE INSULIN.
 GN INS.
 OS Anser anser anser (Western graylag goose), and
 OS Cairina moschata (Muscovy duck).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Anseriformes; Anatidae; Anser.
 OX NCBI_TaxID=8844, 8855;
 RN [1]
 RP SEQUENCE.
 RC SPECIES=A. anser;
 RA Xu Y., Lin N., Zhang Y., Zhang Y.;
 RT "Isolation and sequence determination of goose insulin.";
 RL Xue Tongbao 28:966-968(1983).
 RN [2]
 RP SEQUENCE.
 RC SPECIES=C. moschata;
 RX MEDLINE=96321314; PubMed=8759296;
 RA Chevalier B., Anglade P., Derouet M., Molle D., Simon J.;
 RT "Isolation and characterization of Muscovy (*Cairina moschata*) duck insulin.";


```

RL Comp. Biochem. Physiol. 114B:19-26(1996).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR PIR: JC0007; INGS.
DR HSP: P01308; 1HIS.
DR InterPro: IPR000739; -.
DR Pfam: PF00049; Insulin; 1.
DR PRINTS: PR00276; INSULIN.
DR PRINTS: PR00277; INSULINB.
DR PROSITE: PS00262; INSULIN; 1.
KM Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31
FT CHAIN 31 31 INSULIN A CHAIN.
FT DISULFID 7- 37 INTERCHAIN.
FT DISULFID 19 50 INTERCHAIN.
FT DISULFID 36 41
SQ SEQUENCE 51 AA; 5716 MW; 976FAED8C68386D CRC64;

Query Match 92.3%; Score 36; DB 1; Length 51;
Best Local Similarity 87.5%; Pred. No. 0.42;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HLKSHLV 8
Db 5 HLCGSHLV 12

RESULT 10
INS_BALBO STANDARD; PRT; 51 AA.
ID INS_BALBO
AC P01314;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE INSULIN.
GN INS.
OS Balaenoptera borealis (Sei whale).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;
OC Balaenopteridae; Balaenoptera.
OX NCBI_TaxID=9768;
OX [1]
RN RP SEQUENCE.
RA Ishihara Y., Saito T., Ito Y., Fujino M.;
RT "Structure of sperm- and sei-whale insulins and their breakdown by
RT whale pepsin."
RL Nature 181:1468-1469(1958).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR PIR: A01582; INWHIS.
DR HSP: P01315; 9INS.
DR InterPro: IPR000739; -.
DR Pfam: PF00049; Insulin; 1.
DR PRINTS: PR00276; INSULIN.
DR PRINTS: PR00277; INSULINB.
DR PROSITE: PS00262; INSULIN; 1.
KM Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31

```

```

FT CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 7 37 INTERCHAIN.
FT DISULFID 19 50 INTERCHAIN.
FT DISULFID 36 41
SQ SEQUENCE 51 AA; 5723 MW; 9007B50E400A7DDD CRC64;

Query Match 92.3%; Score 36; DB 1; Length 51;
Best Local Similarity 87.5%; Pred. No. 0.42;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HLKSHLV 8
Db 5 HLCGSHLV 12

RESULT 11
INS_BALPH STANDARD; PRT; 51 AA.
ID INS_BALPH
AC P01312;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE INSULIN.
GN INS.
OS Balaenoptera physalus (Finback whale) (Common rorqual), and
OS Physeter catodon (Sperm whale) (Physeter macrocephalus).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;
OC Balaenopteridae; Balaenoptera.
OX NCBI_TaxID=9770, 9755;
OX [1]
RN RP PARTIAL SEQUENCE.
RC SPECIES-B.physalus;
RA Hama H., Titani K., Sakaki S., Narita K.;
RT "The amino acid sequence in fin-whale insulin."
RL J. Biochem. 56:285-293(1964).
RN [2]
RN RP SEQUENCE.
RC SPECIES-P.catodon;
RA Ishihara Y., Saito T., Ito Y., Fujino M.;
RT "Structure of sperm- and sei-whale insulins and their breakdown by
RT whale pepsin."
RL Nature 181:1468-1469(1958).
RN [3]
RN RP SEQUENCE.
RC SPECIES-P.catodon;
RA Harris J.I., Sanger F., Naughton M.A.;
RT "Species differences in insulin."
RL Arch. Biochem. Biophys. 65:427-438(1956).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR PIR: A91918; INWHF.
DR PIR: A93142; INWHP.
DR HSP: P01315; 6INS.
DR InterPro: IPR000739; -.
DR Pfam: PF00049; Insulin; 1.
DR PRINTS: PR00276; INSULIN.
DR PRINTS: PR00277; INSULINB.
DR PROSITE: PS00262; INSULIN; 1.
KM Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 7 37 INTERCHAIN.
FT DISULFID 19 50 INTERCHAIN.
FT DISULFID 36 41

```

SO SEQUENCE 51 AA: 5766 MW: 9007B514691A7CDD CRC64;

Query Match 92.3%; Score 36; DB 1; Length 51;
Best Local Similarity 87.5%; Pred. No. 0.42;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HLXGSHLV 8
DB 5 HLCGSHLV 12

RESULT 12

INS_CAMDR 12
ID INS_CAMDR STANDARD; PRT; 51 AA.

AC P01320;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE INSULIN.

OS Camelus dromedarius (Dromedary) (Arabian camel).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.
OX NCBI_TaxID=9838;
RN [1]
RA SEQUENCE.

RT "The isolation and characterization of insulin of camel (Camelus dromedarius).";
RL J. Fac. Med. Baghdad 14:16-28(1972).

CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.

CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO DISULFIDE BONDS.

CC -1- SUBCELLULAR LOCATION: SECRETED.

CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.

DR PIR: A92782; INCM.

DR HSSP: P01317; 2INS.

DR InterPro: IPR000739; -

DR Pfam: PF00049; Insulin; 1.

DR PRINTS: PR00276; INSULIN.

DR PROSITE: PS00262; INSULIN; 1.

KW Insulin family; Hormone; Glucose metabolism.

FT CHAIN 1 30 INSULIN B CHAIN.

FT NON_CONS 30 31

FT CHAIN 31 51 INSULIN A CHAIN.

FT DISULFID 7 37 INTERCHAIN.

FT DISULFID 19 50 INTERCHAIN.

FT DISULFID 36 41

SO SEQUENCE 51 AA: 5693 MW: 901E88BA085A7DDD CRC64;

Query Match 92.3%; Score 36; DB 1; Length 51;
Best Local Similarity 87.5%; Pred. No. 0.42;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HLXGSHLV 8
DB 5 HLCGSHLV 12

RESULT 13

INS_CAPHI 13
ID INS_CAPHI STANDARD; PRT; 51 AA.

AC P01319;
DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE INSULIN.

GN INS.

OS Capra hircus (Goat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

OC Bovidae; Caprinae; Capra.

OX NCBI_TaxID=9925;
RN [1]
RA SEQUENCE.

RX MEDLINE=66160119; PubMed=5949593;

RT Smith L.F.;

RL "Species variation in the amino acid sequence of insulin.";
Am. J. Med. 40:662-666(1966).

CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.

CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO DISULFIDE BONDS.

CC -1- SUBCELLULAR LOCATION: SECRETED.

CC PIR: A01586; INGT.

DR HSSP: P01315; 9INS.

DR InterPro: IPR000739; -

DR Pfam: PF00049; Insulin; 1.

DR PRINTS: PR00276; INSULIN.

DR PROSITE: PS00262; INSULIN; 1.

KW Insulin family; Hormone; Glucose metabolism.

FT CHAIN 1 30 INSULIN B CHAIN.

FT NON_CONS 30 31

FT CHAIN 31 51 INSULIN A CHAIN.

FT DISULFID 7 37 INTERCHAIN.

FT DISULFID 19 50 INTERCHAIN.

FT DISULFID 36 41

SO SEQUENCE 51 AA: 5692 MW: 9007B50CBA4E7DDD CRC64;

Query Match 92.3%; Score 36; DB 1; Length 51;
Best Local Similarity 87.5%; Pred. No. 0.42;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HLXGSHLV 8
DB 5 HLCGSHLV 12

RESULT 14

INS_CHIBR 14
ID INS_CHIBR STANDARD; PRT; 51 AA.

AC P01327;
DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE INSULIN.

OS Chinchilla brevicaudata (Chinchilla).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Hystriognathi; Chinchillidae;
OC Chinchilla.

OX NCBI_TaxID=10152;
RN [1]
RA SEQUENCE.

RX MEDLINE=76022416; PubMed=1175610;

RT Wood S.P., Blundell T.L., Wolimer A., Lazarus N.R., Neville R.W.J.;

RT "The relation of conformation and association of insulin to receptor binding; X-ray and circular-dichroism studies on bovine and

RT hystriocomorph insulins.";

RL Eur. J. Biochem. 55:531-542(1975).

CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE

CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.

CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO

DISULFIDE BONDS.

CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR PIR: A01593; INCB.
DR HSSP: P01308; INCB.
DR InterPro: IPR000739; -.
DR Pfam: PF00049; Insulin; 1.
DR PRINTS: PR00276; INSULIN.
DR PRINTS: PR00277; INSULIN.
DR PROSITE: PS00262; INSULIN; 1.
KM Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 7 37 INTERCHAIN (BY SIMILARITY).
FT DISULFID 19 50 INTERCHAIN (BY SIMILARITY).
FT DISULFID 36 41 BY SIMILARITY.
SQ SEQUENCE 51 AA: 5741 MW: 87EC904691A78A0 CRC64;

Query Match 92.3%; Score 36; DB 1; Length 51;
Best Local Similarity 87.5%; Pred. No. 0.42;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 HLXGSHLV 8
|||
Db 5 HLXGSHLV 12

RESULT 15
INS_DIDMA STANDARD; PRT; 51 AA.
AC P18109;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE INSULIN.
GN INS.
OS Didelphis marsupialis virginiana (North American opossum).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Didelphimorphia; Didelphidae; Didelphis.
OX NCBI_Taxid=9267;
RN [1]
RP SEQUENCE.
RC TISSUE=Pancreas;
RX MEDLINE=90160042; PubMed=2695899;
RA Yu J.-H., Eng J., Rattan S., Yalow R.S.;
RT "Opossum Insulin, glucagon and pancreatic polypeptide: amino acid
sequences";
RT Peptides 10:1195-1197(1989).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC PIR: J00362; J00362.
DR PIR: J00363; J00363.
DR HSSP: P01317; 2INS.
DR InterPro: IPR000739; -.
DR Pfam: PF00049; Insulin; 1.
DR PRINTS: PR00276; INSULIN.
DR PRINTS: PR00277; INSULIN.
DR PROSITE: PS00262; INSULIN; 1.
KM Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 7 37 INTERCHAIN.
FT DISULFID 19 50 INTERCHAIN.
FT DISULFID 36 41

SQ SEQUENCE 51 AA: 5732 MW: 9007B8BAE4BDEEDD CRC64;

Query Match 92.3%; Score 36; DB 1; Length 51;
Best Local Similarity 87.5%; Pred. No. 0.42;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 HLXGSHLV 8
|||
Db 5 HLXGSHLV 12

Search completed: September 6, 2001, 16:51:08
Job time: 812 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:45:51 : Search time 134.15 seconds
(without alignments)
4.543 Million cell updates/sec

Title: US-09-603-713-22

Perfect score: 39

Sequence: 1 HLKGSRLV 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	92.3	50	1	INTO2
2	36	92.3	50	1	INTFIS
3	36	92.3	50	1	INTFIS
4	36	92.3	50	1	INTFIS
5	36	92.3	50	1	INTFIS
6	36	92.3	50	1	INTFIS
7	36	92.3	50	1	INTFIS
8	36	92.3	50	1	INTFIS
9	36	92.3	50	1	INTFIS
10	36	92.3	50	1	INTFIS
11	36	92.3	50	1	INTFIS
12	36	92.3	50	1	INTFIS
13	36	92.3	50	1	INTFIS
14	36	92.3	50	1	INTFIS
15	36	92.3	50	1	INTFIS
16	36	92.3	50	1	INTFIS
17	36	92.3	50	1	INTFIS
18	36	92.3	50	1	INTFIS
19	36	92.3	50	1	INTFIS
20	36	92.3	50	1	INTFIS
21	36	92.3	50	1	INTFIS
22	36	92.3	50	1	INTFIS
23	36	92.3	50	1	INTFIS
24	36	92.3	50	1	INTFIS
25	36	92.3	50	1	INTFIS
26	36	92.3	50	1	INTFIS
27	36	92.3	50	1	INTFIS
28	36	92.3	50	1	INTFIS
29	36	92.3	50	1	INTFIS

30	36	92.3	52	1	INCXA	insulin - alligato
31	36	92.3	52	2	S44469	insulin II - North
32	36	92.3	52	2	S44470	insulin I2 - North
33	36	92.3	52	2	S15426	insulin - bowlin
34	36	92.3	54	1	INDF	insulin - splny do
35	36	92.3	57	1	INLMS	insulin - sea lamp
36	36	92.3	68	1	IPRYM	insulin precursor
37	36	92.3	77	1	INSH	insulin precursor
38	36	92.3	81	1	IPDK	insulin precursor
39	36	92.3	84	1	IPPG	insulin precursor
40	36	92.3	86	1	IPHO	insulin precursor
41	36	92.3	96	2	PC7082	epidermal growth f
42	36	92.3	103	2	I51221	insulin precursor
43	36	92.3	105	1	IPBO	insulin precursor
44	36	92.3	105	1	IPON	insulin I precursor
45	36	92.3	106	1	IPXL1	insulin I precursor

ALIGNMENTS

RESULT 1
INTO2
Insulin 2 - loadfish (tentative sequence)
C:Species: Batrachoididae gen. sp. (loadfish)
C:Date: 24-Apr-1984 #sequence_revision 30-Sep-1988 #text_change 31-Mar-2000
C:Accession: A01605
R:Smith, L.F.
Am. J. Med. 40, 662-666, 1966
A:Title: Species variation in the amino acid sequence of Insulin.
A:Reference number: A90029; MUID:66160119
A:Accession: A01605
A:Molecule type: protein
A:Residues: 1-29;30-50 <SMI>
C:Superfamily: Insulin
C:Keywords: hormone; pancreas
F:1-29/Domain: Insulin chain B #status experimental <BCH>
F:1-29,30-50/Product: Insulin #status experimental <MAT>
F:30-50/Domain: Insulin chain A #status experimental <ACH>
F:8-36,20-49,35-40/Disulfide bonds: #status predicted

Query Match 92.3%; Score 36; DB 1; Length 50;
Best Local Similarity 87.5%; Pred. No. 0.58;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HLKGSRLV 8
DB 6 HLKGSRLV 13

RESULT 2
INTFIS
Insulin - shorthorn sculpin
C:Species: Myoxocephalus scorpius (shorthorn sculpin, daddy sculpin)
C:Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 16-Jul-1999
R:Cutfield, J.F.; Cutfield, S.M.; Carne, A.; Emdin, S.O.; Falkner, S.
Eur. J. Biochem. 158, 117-123, 1986
A:Title: The isolation, purification and amino-acid sequence of Insulin from the teleost
A:Reference number: A91169; MUID:86274667
A:Accession: A25061
A:Molecule type: protein
A:Residues: 1-29;30-50 <CVT>
C:Superfamily: Insulin
C:Keywords: hormone; pancreatic islet
F:1-29/Domain: Insulin chain B #status experimental <BCH>
F:1-29,30-50/Product: Insulin #status experimental <MAT>
F:30-50/Domain: Insulin chain A #status experimental <ACH>
F:7-36,19-49,35-40/Disulfide bonds: #status predicted

Query Match 92.3%; Score 36; DB 1; Length 50;

Best Local Similarity 87.5%; Pred. No. 0.58;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HLXGSHLV 8
DB 5 HLXGSHLV 12

RESULT 3

INON
insulin [validated] - pink salmon
C:Species: Oncorhynchus gorbuscha (pink salmon)
C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 11-May-2000
C:Accession: S02203; A60523; S02204
R:Rusakov, Y.I.; Karasev, V.S.; Pertseva, M.N.; Pankov, Y.A.
Biochemistry (N.Y.) 52, 211-217, 1987
A:Title: Amino acid sequence of humpback salmon (Oncorhynchus gorbuscha) insulin.
A:Reference number: S02203
A:Accession: S02203
A:Molecule type: protein
A:Residues: 1-29;30-50 <RUS>
A:Note: this paper is a translation of the Russian paper published in Biokhimiya (1987)
R:Rusakov, Y.I.; Karasev, V.S.; Bondareva, V.M.; Pertseva, M.N.; Pankov, Y.A.
Comp. Biochem. Physiol. B 95, 477-482, 1990
A:Title: Isolation, primary structure, and biological and immunological properties of p
A:Reference number: A60523; M01D:90235564
A:Accession: A60523
A:Molecule type: protein
A:Residues: 1-29;30-50 <RU2>
C:Superfamily: insulin
C:Keywords: hormone; pancreas
F:1-29/Domain: insulin chain B #status experimental <BCH>
F:30-50/Domain: insulin #status experimental <MAT>
F:7-36,19-49,35-40/Disulfide bonds: #status predicted

Query Match
Best Local Similarity 92.3%; Score 36; DB 1; Length 50;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HLXGSHLV 8
DB 5 HLXGSHLV 12

RESULT 4

INONC
insulin - coho salmon
C:Species: Oncorhynchus kisutch (coho salmon)
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 16-Jul-1999
C:Accession: A60322
R:Pilsetskaya, E.; Pollock, H.G.; Rouse, J.B.; Hamilton, J.W.; Kimmel, J.R.; Gorman, A.
Regul. Pept. 11, 105-116, 1985
A:Title: Characterization of coho salmon (Oncorhynchus kisutch) insulin.
A:Reference number: A60322; M01D:85299536
A:Accession: A60322
A:Molecule type: protein
A:Residues: 1-29;30-50 <PLU>
C:Superfamily: insulin
C:Keywords: hormone; pancreas
F:1-29/Domain: insulin chain B #status experimental <BCH>
F:1-29,30-50/Product: insulin #status experimental <MAT>
F:30-50/Domain: insulin chain A #status experimental <ACH>
F:7-36,19-49,35-40/Disulfide bonds: #status predicted

Query Match
Best Local Similarity 92.3%; Score 36; DB 1; Length 50;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HLXGSHLV 8
DB 5 HLXGSHLV 12

DB 5 HLXGSHLV 12

RESULT 5

INBNZ
insulin - skipjack tuna (tentative sequence)
C:Species: Euthynnus pelamis, Katsuwonus pelamis (skipjack tuna)
C:Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change 31-Mar-2000
C:Accession: A01607; A91916
R:Kotaki, A.
J. Biochem. 53, 61-70, 1963
A:Title: Studies on insulin. V. On the structure of the glycy chain of bonito insul
A:Reference number: A91917
A:Accession: A01607
A:Molecule type: protein
A:Residues: 1-29 <KOT1>
R:Kotaki, A.
J. Biochem. 51, 301-309, 1962
A:Title: Studies on insulin. III. On the structure of the alanyl chain of bonito insu
A:Reference number: A91916
A:Accession: A91916
A:Molecule type: protein
A:Residues: 30-50 <KOT2>
C:Superfamily: insulin
C:Keywords: hormone; pancreas
F:1-29/Domain: insulin chain B #status experimental <BCH>
F:1-29,30-50/Product: insulin #status experimental <MAT>
F:30-50/Domain: insulin chain A #status experimental <ACH>
F:7-36,19-49,35-40/Disulfide bonds: #status experimental

Query Match
Best Local Similarity 92.3%; Score 36; DB 1; Length 50;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HLXGSHLV 8
DB 5 HLXGSHLV 12

RESULT 6

INMWP
insulin - sperm whale
C:Species: Physeter catodon (sperm whale)
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 16-Jul-1999
C:Accession: A93142; A90082
R:Ishihara, Y.; Saito, T.; Ito, Y.; Fujino, M.
Nature 181, 1468-1469, 1958
A:Title: Structure of sperm- and sei-whale insulins and their breakdown by whale pep
A:Reference number: A93142
A:Accession: A93142
A:Molecule type: protein
A:Residues: 1-50;31-51 <LSH>
R:Harris, J.L.; Sanger, F.; Naughton, M.A.
Arch. Biochem. Biophys. 65, 427-428, 1956
A:Title: Species differences in insulin.
A:Reference number: A90082
A:Accession: A90082
A:Molecule type: protein
A:Residues: 1-30;31-51 <HAR>
C:Superfamily: insulin
C:Keywords: hormone; pancreas
F:1-30/Domain: insulin chain B #status experimental <BCH>
F:1-30,31-51/Product: insulin #status experimental <MAT>
F:31-51/Domain: insulin chain A #status experimental <ACH>
F:7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match
Best Local Similarity 92.3%; Score 36; DB 1; Length 51;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HLXGSHLV 8
DB 5 HLXGSHLV 12

Db 5 HLXGSHLV 12

RESULT 7

INMNF

Insulin - finback whale (tentative sequence)
C:Species: Balaenoptera physalus (finback whale, common roqual)
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 16-Jul-1999
C:Accession: A91918
R:Hama, H.; Titani, K.; Sakaki, S.; Narita, K.
J. Biochem. 56, 285-293, 1964
A:Title: The amino acid sequence in fin-whale insulin.
A:Reference number: A91918
A:Accession: A91918
A:Molecule type: protein
A:Residues: 1-30:31-51 <HAM>
C:Superfamily: Insulin
C:Keywords: hormone; pancreas
F:1-30/Domain: Insulin chain B #status experimental <BCH>
F:1-30,31-51/Product: Insulin #status experimental <MAT>
F:31-51/Domain: Insulin chain A #status experimental <ACH>
F:7-37,19-50,36-41/Dissulfide bonds: #status predicted

Query Match 92.3%; Score 36; DB 1; Length 51;
Best Local Similarity 87.5%; Pred. No. 0.59;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HLXGSHLV 8
Db 5 HLXGSHLV 12

RESULT 8

INMHS

Insulin - sei whale
C:Species: Balaenoptera borealis (sei whale)
C:Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change 16-Jul-1999
C:Accession: A01582
R:Ishihara, Y.; Salto, T.; Ito, Y.; Fujino, M.
Nature 181, 1468-1469, 1958
A:Title: Structure of sperm- and sei-whale insulins and their breakdown by whale pepsin.
A:Reference number: A93142
A:Accession: A01582
A:Molecule type: protein
A:Residues: 1-30:31-51 <ISH>
C:Superfamily: Insulin
C:Keywords: hormone; pancreas
F:1-30/Domain: Insulin chain B #status experimental <BCH>
F:1-30,31-51/Product: Insulin #status experimental <MAT>
F:31-51/Domain: Insulin chain A #status experimental <ACH>
F:7-37,19-50,36-41/Dissulfide bonds: #status predicted

Query Match 92.3%; Score 36; DB 1; Length 51;
Best Local Similarity 87.5%; Pred. No. 0.59;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HLXGSHLV 8
Db 5 HLXGSHLV 12

RESULT 9

INMEL

Insulin - elephant
C:Species: Elephantidae gen. sp. (elephant)
C:Date: 24-Apr-1984 #sequence_revision 30-Sep-1988 #text_change 16-Jul-1999
C:Accession: A01584
R:Smith, L.F.
Am. J. Med. 40, 662-666, 1966
A:Title: Species variation in the amino acid sequence of insulin.

A:Reference number: A90029; MOID:66160119
A:Accession: A01584
A:Molecule type: protein
A:Residues: 1-30:31-51 <SMR>
A:Note: the species of elephant is not given, but it is most probably the Indian elep

C:Superfamily: Insulin
C:Keywords: hormone; pancreas
F:1-30/Domain: Insulin chain B #status experimental <BCH>
F:1-30,31-51/Product: Insulin #status experimental <MAT>
F:31-51/Domain: Insulin chain A #status experimental <ACH>
F:7-37,19-50,36-41/Dissulfide bonds: #status predicted

Query Match 92.3%; Score 36; DB 1; Length 51;
Best Local Similarity 87.5%; Pred. No. 0.59;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HLXGSHLV 8
Db 5 HLXGSHLV 12

RESULT 10

INMCT

Insulin - goat
C:Species: Capra aegagrus hircus (domestic goat)
C:Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change 16-Jul-1999
C:Accession: A01586
R:Smith, L.F.
Am. J. Med. 40, 662-666, 1966
A:Title: Species variation in the amino acid sequence of insulin.
A:Reference number: A90029; MOID:66160119
A:Accession: A01586
A:Molecule type: protein
A:Residues: 1-30:31-51 <SMT>
C:Superfamily: Insulin
C:Keywords: hormone; pancreas
F:1-30/Domain: Insulin chain B #status experimental <BCH>
F:1-30,31-51/Product: Insulin #status experimental <MAT>
F:31-51/Domain: Insulin chain A #status experimental <ACH>
F:7-37,19-50,36-41/Dissulfide bonds: #status predicted

Query Match 92.3%; Score 36; DB 1; Length 51;
Best Local Similarity 87.5%; Pred. No. 0.59;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HLXGSHLV 8
Db 5 HLXGSHLV 12

RESULT 11

INCKA

Insulin - Arabian camel (tentative sequence)
C:Species: Camelus dromedarius (Arabian camel)
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 16-Jul-1999
C:Accession: A92782
R:Danho, W.O.
J. Fac. Med. Baghdad 14, 16-28, 1972
A:Title: The isolation and characterization of insulin of camel (Camelus dromedarius)
A:Reference number: A92782
A:Accession: A92782
A:Molecule type: protein
A:Residues: 1-30:31-51 <DAN>
C:Superfamily: Insulin
C:Keywords: hormone; pancreas
F:1-30/Domain: Insulin chain B #status experimental <BCH>
F:1-30,31-51/Product: Insulin #status experimental <MAT>
F:31-51/Domain: Insulin chain A #status experimental <ACH>
F:7-37,19-50,36-41/Dissulfide bonds: #status predicted

Query Match 92.3%; Score 36; DB 1; Length 51;
Best Local Similarity 87.5%; Pred. No. 0.59;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HLXGSHLV 8
DB 5 HLCGSHLV 12

RESULT 12

INCR
insulin - cat
C:Species: Felis silvestris catus (domestic cat)
C:Date: 30-Jun-1987 #sequence_revision 30-Sep-1998 #text_change 16-Jul-1999
C:Accession: A01588
R:Hallden, G.; Gafvelin, G.; Mutt, V.; Jornvall, H.
Arch. Biochem. Biophys. 247, 20-27, 1996
A:Title: Characterization of cat insulin.
A:Reference number: A01588; MUID:86214076
A:Accession: A01588
A:Molecule type: protein
A:Residues: 1-30;31-51 <HML>
C:Superfamily: insulin
C:Keywords: hormone; pancreas
F:1-30/Domain: insulin chain B #status experimental <BCH>
F:1-30/31-51/Product: insulin #status experimental <MAT>
F:7-37/19-50/36-41/Disulfide bonds: #status predicted

Query Match 92.3%; Score 36; DB 1; Length 51;
Best Local Similarity 87.5%; Pred. No. 0.59;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HLXGSHLV 8
DB 5 HLCGSHLV 12

RESULT 13

INHY
insulin - hamster
C:Species: Cricetinae gen. sp. (hamster)
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 16-Jul-1999
C:Accession: A91456
R:Neelon, F.A.; Delcher, H.K.; Steinman, H.; Lebovitz, H.E.
Fed. Proc. 32, 300, 1973
A:Title: Structure of hamster insulin: comparison with a tumor insulin.
A:Reference number: A91456
A:Accession: A91456
A:Molecule type: protein
A:Residues: 1-30;31-51 <NEB>
C:Superfamily: insulin
C:Keywords: hormone; pancreas
F:1-30/Domain: insulin chain B #status experimental <BCH>
F:1-30/31-51/Product: insulin #status experimental <MAT>
F:7-37/19-50/36-41/Disulfide bonds: #status predicted

Query Match 92.3%; Score 36; DB 1; Length 51;
Best Local Similarity 87.5%; Pred. No. 0.59;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HLXGSHLV 8
DB 5 HLCGSHLV 12

RESULT 14

INMSSP
insulin - Egyptian spiny mouse (tentative sequence)
C:Species: Acomys cahirinus (Egyptian spiny mouse)

C:Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change 31-Mar-2000
C:Accession: A01591
R:Buenzli, H.F.; Humbel, R.E.
Hoppe-Seyler's Z. Physiol. Chem. 353, 444-450, 1972
A:Title: Isolation and partial structural analysis of insulin from mouse (Mus musculus)
A:Reference number: A01591; MUID:72189454
A:Contents: composition
A:Accession: A01591
A:Molecule type: protein
A:Residues: 1-30;31-51 <BUPE>
C:Superfamily: insulin
C:Keywords: hormone; pancreas
F:1-30/Domain: insulin chain B #status predicted <BCH>
F:1-30/31-51/Product: insulin #status predicted <MAT>
F:7-37/19-50/36-41/Disulfide bonds: #status predicted

Query Match 92.3%; Score 36; DB 1; Length 51;
Best Local Similarity 87.5%; Pred. No. 0.59;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HLXGSHLV 8
DB 5 HLCGSHLV 12

RESULT 15

INCB
insulin - Chinchilla brevicaudata
C:Species: Chinchilla brevicaudata, Chinchilla lanigera brevicaudata
C:Date: 30-Sep-1979 #sequence_revision 30-Sep-1988 #text_change 16-Jul-1999
C:Accession: A01593
R:Wood, S.P.; Blundell, T.L.; Wollmer, A.; Lazarus, N.R.; Neville, R.W.J.
Eur. J. Biochem. 55, 531-542, 1975
A:Title: The relation of conformation and association of insulin to receptor binding;
A:Reference number: A01593; MUID:76022416
A:Accession: A01593
A:Molecule type: protein
A:Residues: 1-30;31-51 <WOO>
C:Superfamily: insulin
C:Keywords: hormone; pancreas
F:1-30/Domain: insulin chain B #status experimental <BCH>
F:1-30/31-51/Product: insulin #status experimental <MAT>
F:7-37/19-50/36-41/Disulfide bonds: #status predicted

Query Match 92.3%; Score 36; DB 1; Length 51;
Best Local Similarity 87.5%; Pred. No. 0.59;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HLXGSHLV 8
DB 5 HLCGSHLV 12

Search completed: September 6, 2001, 16:45:51
Job time: 495 sec

Fri, Sep 7 10:57:46 2001

us-09-603-713-22.rpt

Page 5

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:39:37 : Search time 113.12 Seconds
(without alignments)
1.456 Million cell updates/sec

Title: US-09-603-713-22

Perfect score: 39

Sequence: 1 HLKSHLV 8

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfillsl.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	92.3	22	1	US-08-158-245-2
2	36	92.3	27	1	US-08-304-070-3
3	36	92.3	29	1	US-08-304-070-4
4	36	92.3	29	1	US-08-430-273-2
5	36	92.3	29	2	US-08-992-676-3
6	36	92.3	29	6	5164366-4
7	36	92.3	30	1	US-08-212-696-2
8	36	92.3	30	1	US-08-233-617-2
9	36	92.3	30	1	US-08-160-376A-2
10	36	92.3	30	1	US-08-304-070-2
11	36	92.3	30	1	US-08-285-661-2
12	36	92.3	30	1	US-08-301-838-2
13	36	92.3	30	1	US-08-389-487-5
14	36	92.3	30	1	US-08-342-931-2
15	36	92.3	30	1	US-08-400-256-2
16	36	92.3	30	2	US-08-508-664-7
17	36	92.3	30	2	US-08-353-476-86
18	36	92.3	30	2	US-08-353-476-86
19	36	92.3	30	2	US-08-484-219-6
20	36	92.3	30	2	US-08-929-587-2
21	36	92.3	30	2	US-08-992-676-2
22	36	92.3	30	2	US-09-134-836-2
23	36	92.3	30	3	US-08-975-365-2
24	36	92.3	30	3	US-08-750-391-2
25	36	92.3	30	3	US-08-750-391-4
26	36	92.3	30	3	US-08-622-046B-20
27	36	92.3	30	4	US-09-255-66B-1

28	36	92.3	30	4	US-09-099-307-2	Sequence 2, Appl1
29	36	92.3	30	4	US-09-099-307-3	Sequence 3, Appl1
30	36	92.3	30	4	US-09-099-307-4	Sequence 4, Appl1
31	36	92.3	30	4	US-09-099-307-5	Sequence 5, Appl1
32	36	92.3	30	4	US-09-099-307-10	Sequence 10, Appl1
33	36	92.3	30	4	US-08-900-574-2	Sequence 2, Appl1
34	36	92.3	30	4	US-08-932-082-2	Sequence 3, Appl1
35	36	92.3	30	5	PCT-US94-04179-3	Sequence 4, Appl1
36	36	92.3	30	5	PCT-US94-04179-4	Sequence 5, Appl1
37	36	92.3	30	6	5164366-6	Patent No. 5164366
38	36	92.3	30	6	5169865-5	Patent No. 5169865
39	36	92.3	30	6	5116465-51	Patent No. 5116465
40	36	92.3	30	6	5116465-52	Patent No. 5116465
41	36	92.3	31	1	US-08-389-487-10	Sequence 10, Appl1
42	36	92.3	31	2	US-08-992-676-5	Sequence 5, Appl1
43	36	92.3	32	1	US-08-087-831-1	Sequence 1, Appl1
44	36	92.3	32	1	US-08-389-487-9	Sequence 9, Appl1
45	36	92.3	32	1	US-08-466-945-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-08-158-245-2
: Sequence 2, Application US/08158245
: Patent No. 5442043
: GENERAL INFORMATION:
: APPLICANT: FUKUTA, Makoto
: APPLICANT: IINUMA, Satoshi
: APPLICANT: OKADA, Hiroaki
: TITLE OF INVENTION: PEPTIDE CONJUGATE
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: WEGNER, CANTOR, MUELLER & PLAYER
: STREET: 1233 20th Street, N.W., Suite 300
: CITY: Washington
: STATE: D.C.
: COUNTRY: U.S.A.
: ZIP: 20036-8218
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/158,245
: FILING DATE:
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 318031-1992
: FILING DATE: 27-NOV-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Cantor, Herbert I.
: REGISTRATION NUMBER: 24,392
: REFERENCE/DOCKET NUMBER: P-8700-24068
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 887-0400
: TELEFAX: (202) 835-0605
: TELEX: 440706 and 248394
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 22 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: US-08-158-245-2

Query Match 92.3% Score 36; DB 1; Length 22;
Best Local Similarity 87.5% Pred. No. 0.35; 1; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 1;

OY 1 HLXGSHLV 8
DB 5 HLXGSHLV 12

RESULT 2

US-08-304-070-3
Sequence 3, Application US/08304070
Patent No. 5547929
GENERAL INFORMATION:
APPLICANT: Anderson Jr., James H.
APPLICANT: De Felippis, Michael R.
APPLICANT: Frank, Bruce H.
APPLICANT: Havel, Henry A.
TITLE OF INVENTION: INSULIN ANALOG FORMULATIONS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company/Patent Division
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana
COUNTRY: United States
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/304,070
FILING DATE: 12-SEP-1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Caltrider, Steven P.
REGISTRATION NUMBER: 36,467
REFERENCE/DOCKET NUMBER: X-9635
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-0757
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 27 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-304-070-3

Query Match 92.3%; Score 36; DB 1; Length 27;
Best Local Similarity 87.5%; Pred. No. 0.44;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HLXGSHLV 8
DB 5 HLXGSHLV 12

RESULT 3

US-08-304-070-4
Sequence 4, Application US/08304070
Patent No. 5547929
GENERAL INFORMATION:
APPLICANT: Anderson Jr., James H.
APPLICANT: De Felippis, Michael R.
APPLICANT: Frank, Bruce H.
APPLICANT: Havel, Henry A.
TITLE OF INVENTION: INSULIN ANALOG FORMULATIONS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company/Patent Division
STREET: Lilly Corporate Center
CITY: Indianapolis

STATE: Indiana
COUNTRY: United States
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/304,070
FILING DATE: 12-SEP-1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Caltrider, Steven P.
REGISTRATION NUMBER: 36,467
REFERENCE/DOCKET NUMBER: X-9635
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-0757
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 29 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-304-070-4

Query Match 92.3%; Score 36; DB 1; Length 29;
Best Local Similarity 87.5%; Pred. No. 0.47;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HLXGSHLV 8
DB 5 HLXGSHLV 12

RESULT 4

US-08-430-273-2
Sequence 2, Application US/08430273
Patent No. 5621073
GENERAL INFORMATION:
APPLICANT: Dickhardt, Rainer
APPLICANT: Unger, Bernhard
APPLICANT: Grafe, Claudia
TITLE OF INVENTION: Process for Obtaining Insulin-Containing
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/430,273
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/991,261
FILING DATE: 16-DEC-1992
APPLICATION NUMBER: DE P 4141794.1
FILING DATE: 18-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE P 4220293.0
FILING DATE: 20-JUN-1992

ATTORNEY/AGENT INFORMATION:
NAME: Einaudi, Carol P.
REGISTRATION NUMBER: 32,220
REFERENCE/DOCKET NUMBER: 02481-1242-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 29 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-430-273-2

Query Match 92.3%; Score 36; DB 1; Length 29;
Best Local Similarity 87.5%; Pred. No. 0.47;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HLXGSHLV 8
|| |||||
DB 5 HLCGSHLV 12

RESULT 5
US-08-992-676-3
Sequence 3, Application US/08992676
Patent No. 5977297
GENERAL INFORMATION:
APPLICANT: OBERMEIER, Rainer
APPLICANT: LUDWIG, Jurgen
APPLICANT: SABEL, Walter
TITLE OF INVENTION: A Process for Isolating Insulin Using
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: FINNEGAN HENDERSON FARABOW GARRETT & DUNNER,
ADDRESS: L.L.P.
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/992,676
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Jensen, Allen R.
REGISTRATION NUMBER: 28,224
REFERENCE/DOCKET NUMBER: 02481.1567-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 29 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
US-08-992-676-3

Query Match 92.3%; Score 36; DB 2; Length 29;

Best Local Similarity 87.5%; Pred. No. 0.47;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HLXGSHLV 8
|| |||||
DB 5 HLCGSHLV 12

RESULT 6
5164366-4
Patent No. 5164366
APPLICANT: BALSCHMIDT, PER, BRANGE, JENS J.V.
TITLE OF INVENTION: HUMAN INSULIN ANALOGUES
NUMBER OF SEQUENCES: 14
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/453,445
FILING DATE: 20-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 416,218
FILING DATE: 2-OCT-1989
APPLICATION NUMBER: 332,697
FILING DATE: 03-APR-1989
SEQ ID NO: 4
LENGTH: 29
5164366-4

Query Match 92.3%; Score 36; DB 6; Length 29;
Best Local Similarity 87.5%; Pred. No. 0.47;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HLXGSHLV 8
|| |||||
DB 5 HLCGSHLV 12

RESULT 7
US-08-212-696-2
Sequence 2, Application US/08212696
Patent No. 5422339
GENERAL INFORMATION:
APPLICANT: George S. Eisenbarth et al.
TITLE OF INVENTION: PEPTIDES HAVING INSULIN
TITLE OF INVENTION: AUTOANTIBODY BUT NOT
TITLE OF INVENTION: INSULIN RECEPTOR BINDING
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
COMPUTER: IBM PS/2 Model 502 or 555X
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: Mordperfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/212,696
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/671,455
FILING DATE: 03/19/91
ATTORNEY/AGENT INFORMATION:
NAME: Freeman, John W.
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 00303/009001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906

TELEX: 200154
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 30
TYPE: amino acid
STRANDEDNESS: N/A
TOPOLOGY: linear
US-08-212-696-2

Query Match 92.3% Score 36; DB 1; Length 30;
Best Local Similarity 87.5% Pred. No. 0.49;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HLXGSHLV 8
|||
Db 5 HLCGSHLV 12

RESULT 8
US-08-233-617-2
Sequence 2, Application US/08233617
Patent No. 5466666
GENERAL INFORMATION:
APPLICANT: Obermeier, Rainer
APPLICANT: Sabel, Walter
APPLICANT: Dell, Peter
APPLICANT: Gelsen, Karl
TITLE OF INVENTION: Amorphous Monospherical Forms of Insulin
TITLE OF INVENTION: Derivatives
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flinnegan, Henderson, Farabow, Garrett &
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,617
FILING DATE: 25-APR-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: P 43 13 702.4
FILING DATE: 27-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Carol P. Einaudi
REGISTRATION NUMBER: 32,220
REFERENCE/DOCKET NUMBER: 02481.1374-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
US-08-233-617-2

Query Match 92.3% Score 36; DB 1; Length 30;
Best Local Similarity 87.5% Pred. No. 0.49;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HLXGSHLV 8
|||
Db 5 HLCGSHLV 12

RESULT 9
US-08-160-376A-2
Sequence 2, Application US/08160376A
Patent No. 5473049
GENERAL INFORMATION:
APPLICANT: Obermeier, Rainer
APPLICANT: Gerl, Martin
APPLICANT: Ludwig, Jurgen
APPLICANT: Sabel, Walter
TITLE OF INVENTION: Process For Obtaining Proinsulin
TITLE OF INVENTION: Possessing Correctly Linked
TITLE OF INVENTION: Cysteine Bridges
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kenneth A. Genoni, Esq.
STREET: Rt. 202-206 No. 5473049th/P.O. Box 2500
CITY: Somerville
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 08876-1258
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
COMPUTER: IBM 386
OPERATING SYSTEM: WINDOWS 3.1
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/160,376A
FILING DATE: December 1, 1993
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GE P 4240420.7
FILING DATE: December 2, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Barbara V. Maurer, Esq.
REGISTRATION NUMBER: 31,287
REFERENCE/DOCKET NUMBER: HOE 92/F 384
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 231-4079
TELEFAX: (908) 231-2255
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 Amino Acids
TYPE: Amino Acid
TOPOLOGY: not relevant
US-08-160-376A-2

Query Match 92.3% Score 36; DB 1; Length 30;
Best Local Similarity 87.5% Pred. No. 0.49;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HLXGSHLV 8
|||
Db 5 HLCGSHLV 12

RESULT 10
US-08-304-070-2
Sequence 2, Application US/08304070
Patent No. 5547929
GENERAL INFORMATION:
APPLICANT: Anderson Jr., James H.
APPLICANT: De Felippis, Michael R.
APPLICANT: Frank, Bruce H.
APPLICANT: Havel, Henry A.
TITLE OF INVENTION: INSULIN ANALOG FORMULATIONS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:

ADDRESSEE: Eli Lilly and Company/Patent Division
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana
COUNTRY: United States
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/304,070
FILING DATE: 12-SEP-1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Caltrider, Steven P.
REGISTRATION NUMBER: 36,467
REFERENCE/DOCKET NUMBER: X-9635
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-0757
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 28
OTHER INFORMATION: /note="Xaa at position 28 of SEQ
OTHER INFORMATION: ID NO:2 is Asp, Lys, Leu, Val, or Ala."
FEATURE:
NAME/KEY: Modified-site
LOCATION: 29
OTHER INFORMATION: /note="Xaa at position 29 of SEQ
OTHER INFORMATION: ID NO:2 is Lys or Pro."
US-08-304-070-2

Query Match 92.3%; Score 36; DB 1; Length 30;
Best Local Similarity 87.5%; Pred. No. 0.49;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 1 HLKSHLV 8
|| |||||
DB 5 HLCGSHLV 12

RESULT 11
US-08-285-661-2
Sequence 2, Application US/08285661
Patent No. 5559094
GENERAL INFORMATION:
APPLICANT: Brems et. al.
TITLE OF INVENTION: Asp Blnsulin Analogs
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
ADDRESS: Patent Division/DKN
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.0 Mb storage
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/285,661

FILING DATE: 530
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Douglas K. No. 5559094man
REGISTRATION NUMBER: 33267
REFERENCE/DOCKET NUMBER: X-9142
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317) 276-2958
TELEFAX: (317) 276-1294
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: polypeptide
FEATURE:
NAME/KEY: Variable site
LOCATION: 13
IDENTIFICATION METHOD:
OTHER INFORMATION: "This amino acid is either Gln or Glu."
US-08-285-661-2

Query Match 92.3%; Score 36; DB 1; Length 30;
Best Local Similarity 87.5%; Pred. No. 0.49;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 1 HLKSHLV 8
|| |||||
DB 5 HLCGSHLV 12

RESULT 12
US-08-301-838-2
Sequence 2, Application US/08301838
Patent No. 5597796
GENERAL INFORMATION:
APPLICANT: Brange, Jens J.V.
TITLE OF INVENTION: TRANSFERMAL INSULIN
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 559779660 No. 55977966disk of No. 55977966th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/301,838
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION NUMBER: US 08/085,473
APPLICATION NUMBER: US 30-JUN-1993
FILING DATE: 30-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/751,836
FILING DATE: 30-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DK 0101/91
FILING DATE: 22-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/DK92/00019
FILING DATE: 22-JAN-1992

ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 331,728
REFERENCE/DOCKET NUMBER: 3447.210-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 7
OTHER INFORMATION: /note- "this disulfide bond is
OTHER INFORMATION: between residue number 7 of this sequence and
OTHER INFORMATION: residue number 7 of SEQ ID NO:1"
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 19
OTHER INFORMATION: /note- "this disulfide bond is
OTHER INFORMATION: between residue number 19 of this sequence and
OTHER INFORMATION: residue number 20 of SEQ ID NO:1"
US-08-301-838-2

Query Match 92.3%; Score 36; DB 1; Length 30;
Best Local Similarity 87.5%; Pred. No. 0.49;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HLXGSHLV 8
11111111
DB 5 HLCGSHLV 12

RESULT 13
US-08-389-487-5
Sequence 5, Application US/08389487
Patent No. 5663291
GENERAL INFORMATION:
APPLICANT: Obermeyer, Rainer
APPLICANT: Gerl, Martin
APPLICANT: Ludwig, Jürgen
APPLICANT: Sabel, Walter
TITLE OF INVENTION: Process for Obtaining Insulin Having
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESS: Dunnet
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/389,487
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Einaudi, Carol P.
REGISTRATION NUMBER: 32,220
REFERENCE/DOCKET NUMBER: 02481.1424-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000

TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-389-487-5

Query Match 92.3%; Score 36; DB 1; Length 30;
Best Local Similarity 87.5%; Pred. No. 0.49;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HLXGSHLV 8
11111111
DB 5 HLCGSHLV 12

RESULT 14
US-08-342-931-2
Sequence 2, Application US/08342931
Patent No. 5693609
GENERAL INFORMATION:
APPLICANT: Baker et. al.
TITLE OF INVENTION: Acylated Insulin Analogs
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
ADDRESSEE: Patent Division/SPC
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/342,931
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Steven P. Caltrider
REGISTRATION NUMBER: 36467
REFERENCE/DOCKET NUMBER: X9720
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317) 276-0757
TELEFAX: (317) 277-1917
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 amino acids
TYPE: amino acid
MOLECULE TYPE: polypeptide
FEATURE:
NAME/KEY: Variable site
LOCATION: 1
IDENTIFICATION METHOD:
OTHER INFORMATION: "Xaa at position 1 of SEQ ID NO:2 is Phe; or acylated Phe
FEATURE:
NAME/KEY: Variable site
LOCATION: 28
IDENTIFICATION METHOD:
OTHER INFORMATION: "Xaa at position 28 of SEQ ID NO:2 is Asp, Lys, Leu, Val,
FEATURE:

NAME/KEY: Variable Site
LOCATION: 29
IDENTIFICATION METHOD:
OTHER INFORMATION: "Xaa at position 29 of SEQ ID NO:2 is Lys, Pro; or acylated L
US-08-342-931-2

Query Match 92.3%; Score 36; DB 1; Length 30;
Best Local Similarity 87.5%; Pred. No. 0.49;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HLXGSHLV 8
|||
Db 5 HLXGSHLV 12

RESULT 15
US-08-400-256-2

; Sequence 2, Application US/08400256
; Patent No. 5750497

; GENERAL INFORMATION:

; APPLICANT: Havelund, Svend

; APPLICANT: Halstrom, John

; APPLICANT: Jonassen, Ib

; APPLICANT: Andersen, Asger Sloth

; APPLICANT: Markussen, Jan

; TITLE OF INVENTION: ACYLATED INSULIN

; NUMBER OF SEQUENCES: 49

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 57504970 No. 5750497disk of No. 5750497th America, Inc.

; STREET: 405 Lexington Avenue, 64th Floor

; CITY: New York

; STATE: New York

; COUNTRY: United States of America

; ZIP: 10174-6401

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/400,256

; FILING DATE: 03-MAR-1995

; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:

; NAME: Lambiris, Elias J.

; REGISTRATION NUMBER: 33,728

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-867-0123

; TELEFAX: 212-878-9655

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 30 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-400-256-2

Query Match 92.3%; Score 36; DB 1; Length 30;
Best Local Similarity 87.5%; Pred. No. 0.49;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HLXGSHLV 8
|||
Db 5 HLXGSHLV 12

Search completed: September 6, 2001, 16:39:37
Job time: 126 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:43:29 : Search time 225.25 Seconds
(without alignments)
2.153 Million cell updates/sec

Title: US-09-603-713-22

Perfect score: 39

Sequence: 1 HLKSHLV 8

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

A.Geneseq_0601:*

- 1: /SID8/gcgdata/geneseq/AA1980.DAT:*
- 2: /SID8/gcgdata/geneseq/AA1981.DAT:*
- 3: /SID8/gcgdata/geneseq/AA1982.DAT:*
- 4: /SID8/gcgdata/geneseq/AA1983.DAT:*
- 5: /SID8/gcgdata/geneseq/AA1984.DAT:*
- 6: /SID8/gcgdata/geneseq/AA1985.DAT:*
- 7: /SID8/gcgdata/geneseq/AA1986.DAT:*
- 8: /SID8/gcgdata/geneseq/AA1987.DAT:*
- 9: /SID8/gcgdata/geneseq/AA1988.DAT:*
- 10: /SID8/gcgdata/geneseq/AA1989.DAT:*
- 11: /SID8/gcgdata/geneseq/AA1990.DAT:*
- 12: /SID8/gcgdata/geneseq/AA1991.DAT:*
- 13: /SID8/gcgdata/geneseq/AA1992.DAT:*
- 14: /SID8/gcgdata/geneseq/AA1993.DAT:*
- 15: /SID8/gcgdata/geneseq/AA1994.DAT:*
- 16: /SID8/gcgdata/geneseq/AA1995.DAT:*
- 17: /SID8/gcgdata/geneseq/AA1996.DAT:*
- 18: /SID8/gcgdata/geneseq/AA1997.DAT:*
- 19: /SID8/gcgdata/geneseq/AA1998.DAT:*
- 20: /SID8/gcgdata/geneseq/AA1999.DAT:*
- 21: /SID8/gcgdata/geneseq/AA2000.DAT:*
- 22: /SID8/gcgdata/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	94.9	8	22	AA66580
2	37	94.9	8	22	AA661342
3	37	94.9	26	22	AA648863
4	37	94.9	30	21	AA610448
5	37	94.9	30	22	AA648853
6	37	94.9	30	22	AA648855
7	37	94.9	30	22	AA648857
8	37	94.9	30	22	AA648859
9	37	94.9	30	22	AA648861
10	36	92.3	12	16	AA682088
11	36	92.3	13	19	AA651790

12	36	92.3	16	15	AA651919	Insulin (B1-B16).
13	36	92.3	22	15	AA66218	Insulin (B1-B22).
14	36	92.3	22	16	AA675413	Insulin B-chain de
15	36	92.3	25	21	AA610773	Insulin beta chain
16	36	92.3	26	17	AA604884	N(epsilomB26)-tetr
17	36	92.3	27	17	AA604883	N(epsilomB27)-tetr
18	36	92.3	28	17	AA604882	N(epsilomB28)-tetr
19	36	92.3	28	19	AA660065	Human insulin B-ch
20	36	92.3	28	19	AA660066	Human insulin B-ch
21	36	92.3	28	19	AA660063	Human insulin B-ch
22	36	92.3	28	19	AA660064	Human insulin B-ch
23	36	92.3	28	19	AA646931	Human insulin B ch
24	36	92.3	28	19	AA646924	Human insulin B ch
25	36	92.3	29	3	AA620139	Sequence of des-Pn
26	36	92.3	29	3	AA620262	Modified Insulin B
27	36	92.3	29	6	AA650834	Sequence of des-Pn
28	36	92.3	29	11	AA608197	Insulin derivative
29	36	92.3	29	11	AA608467	Example of pref. m
30	36	92.3	29	15	AA652573	Insulin B chain fr
31	36	92.3	29	17	AA604886	N(1-carboxytridecy
32	36	92.3	29	17	AA604887	Tetradecanoyl-tetr
33	36	92.3	29	17	AA604889	N(epsilomB29)-tetr
34	36	92.3	29	19	AA670341	Insulin analogue 1
35	36	92.3	29	19	AA670343	Insulin analogue 3
36	36	92.3	29	19	AA663756	Insulin protein fr
37	36	92.3	29	19	AA660068	Human insulin B-ch
38	36	92.3	29	19	AA660059	Human insulin B-ch
39	36	92.3	29	19	AA660060	Human insulin B-ch
40	36	92.3	29	19	AA660061	Human insulin B-ch
41	36	92.3	29	19	AA660062	Human insulin B-ch
42	36	92.3	29	19	AA646929	Human insulin B ch
43	36	92.3	29	19	AA646935	Human insulin B ch
44	36	92.3	29	19	AA641337	Human insulin B ch
45	36	92.3	29	19	AA646922	Insulin B chain an

ALIGNMENTS

RESULT	1
AA66580	standard; Peptide: 8 AA.
ID	AA66580
AC	AA66580;
XX	
DT	12-APR-2001 (first entry)
XX	
DE	Oxidised insulin B-chain peptide #1.
XX	
KW	memapsin 2; nootropic; neuroprotective; amyloid precursor protein;
KW	APP; memapsin 2 inhibitor; Alzheimer's disease; insulin B-chain peptide.
XX	
OS	Synthetic.
XX	
FT	Key
FT	Modified-site
FT	Location/Qualifiers
XX	3
XX	/note- "Cysteic acid"
PN	WO200100665-A2.
PD	
XX	04-JAN-2001.
XX	
PF	27-JUN-2000; 2000WO-US17742.
XX	
PR	28-JUN-1999; 99US-0141363.
PR	30-NOV-1999; 99US-0168060.
PR	25-JAN-2000; 2000US-0177836.
PR	27-JAN-2000; 2000US-0178368.
PR	08-JUN-2000; 2000US-0210292.
XX	
PA	(OKLA-) OKLAHOMA MEDICAL RES FOUND.
PA	(UNIT) UNIV ILLINOIS FOUND.
XX	

PI Tang JUN, Hong L, Ghosh AK;
 XX WPI: 2001-137933/14.
 DR
 XX Novel memapsin 2 inhibitors which bind to active site of memapsin 2
 PT having 2 catalytic aspartic residues and substrate binding cleft, used
 PT to treat Alzheimer's disease by blocking amyloid precursor protein
 PT cleavage
 XX
 PS Example 4; Page 33; 86pp; English.
 CC The present sequence is given in a specification relating to an inhibitor
 CC of catalytically active memapsin 2. The inhibitor binds to the memapsin 2
 CC active site, which is defined by the presence of two catalytic aspartic
 CC residues and a substrate binding cleft. The inhibitor is useful for
 CC the treatment and diagnosis of Alzheimer's disease. It is useful in
 CC screens for individuals with a genetic predisposition to Alzheimer's
 CC disease. The inhibitor is useful as a reagent for specifically binding to
 CC memapsin 2 or memapsin 2 analogues and for aiding in memapsin 2
 CC isolation, purification and characterisation.
 CC
 CC Sequence 8 AA;
 SQ
 Query Match 94.9%; Score 37; DB 22; Length 8;
 Best Local Similarity 100.0%; Pred. No. 3.4e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 HIXGSHLV 8
 DB 1 hlxgshlv 8
 DB
 RESULT 2
 AAB61342
 ID AAB61342 standard; peptide; 8 AA.
 AC AAB61342;
 XX
 XX 02-APR-2001 (first entry)
 DT
 XX Memapsin 2 substrate #1.
 DE
 XX Memapsin 2; catalyst; Alzheimer's.
 KW
 XX Unidentified.
 OS
 XX W0200100663-A2.
 PN
 XX 04-JAN-2001.
 PD
 XX 27-JUN-2000; 2000WO-US17661.
 PF
 XX 28-JUN-1999; 99US-0141363.
 PR 30-NOV-1999; 99US-0168060.
 PR 25-JAN-2000; 2000US-0177836.
 PR 27-JAN-2000; 2000US-0178368.
 PR 08-JUN-2000; 2000US-0210292.
 XX
 PA (OKLA-) OKLAHOMA MEDICAL RES FOUND.
 XX
 XX Tang JUN, Lin X, Koelsch G;
 PI WPI: 2001-102885/11.
 DR
 XX Purified recombinant catalytically active memapsin 2, used to screen
 PT inhibitors of it, which are used to treat and prevent Alzheimer's
 PT disease
 XX
 PS Example 4; Page 33; 86pp; English.
 CC The present invention relates to a purified recombinant
 CC catalytically active memapsin 2. The invention may be used for

CC isolating inhibitors which are used to treat or prevent
 CC Alzheimer's disease. The invention may also be used to screen
 CC for individuals more genetically prone to develop Alzheimer's
 CC disease.
 CC
 CC Sequence 8 AA;
 SQ
 Query Match 94.9%; Score 37; DB 22; Length 8;
 Best Local Similarity 100.0%; Pred. No. 3.4e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 HIXGSHLV 8
 DB 1 hlxgshlv 8
 DB
 RESULT 3
 AAB48863
 ID AAB48863 standard; peptide; 26 AA.
 AC AAB48863;
 XX
 XX 16-MAR-2001 (first entry)
 DT
 XX Mutant human insulin (IA protein) cys-4 B chain.
 DE
 XX Human insulin; insulin activity protein; IA protein;
 KW diabetes mellitus; type 1; type 2; stability;
 KW disulphide bond; cysteine replacement; mutant; mutein.
 OS
 XX Homo sapiens.
 OS Synthetic.
 XX W0200069901-A2.
 PN
 XX 23-NOV-2000.
 PD
 XX 19-MAY-2000; 2000WO-US13764.
 PF
 XX 19-MAY-1999; 99US-0134930.
 PR
 XX (XENC-) XENCOR INC.
 PA
 XX Dabiyat BI;
 PI WPI: 2001-025004/03.
 DR
 XX Non-naturally occurring protein with insulin activity useful for
 PT treating type 1 and type 2 diabetes, comprising amino acid
 PT substitutions as compared to native human insulin and having enhanced
 PT stability
 XX
 XX Claim 10; Fig 3G; 95pp; English.
 CC The invention relates to novel non-naturally occurring mature human
 CC insulin mutants, designated insulin activity (IA) proteins in the
 CC specification (AAB48863-A48876), which have altered properties (e.g.,
 CC enhanced oxidative, alkaline or thermal stability, prolonged shelf-life,
 CC altered ability to oligomerise) compared to wild-type mature human
 CC insulin (AAB48848, AAB48849) but which are still capable of binding to
 CC an insulin receptor. The insulin mutants of the invention have less than
 CC 98% sequence identity to wild-type human insulin, but have a 3D
 CC structure which substantially corresponds to that of wild-type human
 CC insulin. The invention also relates to recombinant nucleic acids encoding
 CC the insulin mutants, expression vectors and host cells comprising mutant
 CC insulin-encoding DNA, the recombinant preparation of an insulin mutant,
 CC and a pharmaceutical composition comprising an insulin mutant of the
 CC invention. The mature human insulin mutants are useful for treating
 CC insulin-responsive conditions and disorders of carbohydrate metabolism
 CC such as type 1 or type 2 diabetes. The nucleic acids encoding the
 CC insulin mutants are used in gene therapy techniques for treating these
 CC conditions. The insulin mutants are more stable compared to wild-type

CC mature human insulin, or have an improved ability to form insulin
 CC hexamers. Sequences AAB4850-B4863 represent the A and B chains of
 CC insulin mutants in which one or more cysteine residue has been
 CC replaced, thus preventing the formation of at least one disulphide bond
 CC and improving stability and activity.

XX Sequence 26 AA;

Query Match 94.9%; Score 37; DB 22; Length 26;

Best Local Similarity 87.5%; Pred. No. 0.32;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HLXGSHLV 8

DB 1 hlxgshlv 8

RESULT 4

AAB10448

ID AAB10448 standard; peptide: 30 AA.

XX AAB10448;

DT 01-DEC-2000 (first entry)

XX Bovine insulin beta-chain peptide SEQ ID NO: 4.

DE Peptide sequencing; fragmentation pattern; amino acid modification;

KW post-translational modification; laundry; cleansing product; proteomic;

KM y-ion; bovine.

XX Bos taurus.

OS WO200043792-A2.

PN 27-JUL-2000.

PD 12-JAN-2000; 2000WO-US00790.

PF 20-JAN-1999; 99US-0116502.

PR 29-SEP-1999; 99US-0136677.

XX (PROC) PROCTER & GAMBLE CO.

PA Keough TW, Youngquist RS;

PI WPI: 2000-543265/49.

DR Determining amino acid sequence of polypeptide by derivatizing the

XX N-terminus of the polypeptide with acidic moieties, analyzing

PT derivatized products using mass spectrometric technique and

PT interpreting the fragmentation pattern -

XX Example 13; Page 30; 30pp; English.

PS This invention describes a novel method for determining the amino acid

CC sequence of a polypeptide comprising derivatizing the N-terminus of the

CC polypeptide or polypeptides with one or more acidic moieties with pK_a of

CC less than 2 when coupled with the polypeptide or polypeptides, analyzing

CC the derivatized products using a mass spectrometric technique to provide

CC a fragmentation pattern free of a- and b-ions and interpreting the

CC fragmentation pattern. The method is used for sequencing wild-type or

CC variant polypeptides. Applications include biological studies,

CC identification of post-translational modifications in proteins,

CC e.g. commercial laundry and cleansing products, designing oligonucleotide

CC probes for gene cloning, rapid characterization of products formed in

CC directed evolution studies, combinatorial chemistry and peptide libraries

CC and proteomics. Derivatization of the polypeptides with acid groups gives

CC almost exclusive y-ion fragmentation and very little a-ion and b-ion

CC 'noise', providing mass spectra which are more easily interpreted. The

CC method is simple, efficient and widely applicable to both wild-type and

CC variant polypeptides. This sequence represents a fragment of the bovine
 CC insulin beta chain which is used to illustrate the method of the
 CC invention.

XX Sequence 30 AA;

Query Match 94.9%; Score 37; DB 21; Length 30;

Best Local Similarity 100.0%; Pred. No. 0.37;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HLXGSHLV 8

DB 5 hlxgshlv 12

RESULT 5

AAB48853

ID AAB48853 standard; peptide: 30 AA.

XX AAB48853;

DT 16-MAR-2001 (first entry)

XX Mutant human insulin (IA protein) cyst77a B chain.

DE Human insulin; insulin activity protein; IA protein;

KW diabetes mellitus; type 1; type 2; stability;

KM disulphide bond; cysteine replacement; mutant; muten.

XX Homo sapiens.

OS Synthetic.

PN WO200069901-A2.

PD 23-NOV-2000.

PF 19-MAY-2000; 2000WO-US13764.

PR 19-MAY-1999; 99US-0134930.

XX (XENC-) XENCOR INC.

PA Dahiyat BI;

PI WPI: 2001-025004/03.

DR Non-naturally occurring protein with insulin activity useful for

XX treating type 1 and type 2 diabetes, comprising amino acid

PT substitutions as compared to native human insulin and having enhanced

PT stability -

XX Claim 10; Fig 3B; 95pp; English.

PS The invention relates to novel non-naturally occurring mature human

CC insulin mutants, designated insulin activity (IA) proteins in the

CC specification (AAB4850-B4876), which have altered properties (e.g.,

CC enhanced oxidative, alkaline or thermal stability, prolonged shelf-life,

CC altered ability to oligomerise) compared to wild-type mature human

CC insulin (AAB4848, AAB4849) but which are still capable of binding to

CC an insulin receptor. The insulin mutants of the invention have less than

CC 98% sequence identity to wild-type human insulin, but have a 3D

CC structure which substantially corresponds to that of wild-type human

CC insulin. The invention also relates to recombinant nucleic acids encoding

CC the insulin mutants, expression vectors and host cells comprising mutant

CC insulin-encoding DNA, the recombinant preparation of an insulin mutant,

CC and a pharmaceutical composition comprising an insulin mutant of the

CC invention. The mature human insulin mutants are useful for treating

CC insulin-responsive conditions and disorders of carbohydrate metabolism

CC such as type 1 or type 2 diabetes. The nucleic acids encoding the

CC insulin mutants are used in gene therapy techniques for treating these

CC conditions. The insulin mutants are more stable compared to wild-type

CC mature human insulin, or have an improved ability to form insulin

CC hexamers. Sequences AAB48850-B48863 represent the A and B chains of
CC insulin mutants in which one or more cysteine residue has been
CC replaced, thus preventing the formation of at least one disulphide bond
CC and improving stability and activity.
XX
SQ Sequence 30 AA;

Query Match 94.9%; Score 37; DB 22; Length 30;
Best Local Similarity 87.5%; Pred. No. 0.37;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HLXGSHLV 8
|||
5 hlygshlv 12

RESULT 6

ID AAB48855 standard; peptide; 30 AA.

AC AAB48855;

DT 16-MAR-2001 (first entry)

DE Mutant human insulin (IA protein) cys77b B chain.

XX Human insulin; insulin activity protein; IA protein;

KW diabetes mellitus; type 1; type 2; stability;

KM disulphide bond; cysteine replacement; mutant; mutein.

XX Homo sapiens.

OS Synthetic.

PN WO200069901-A2.

PD 23-NOV-2000.

PF 19-MAY-2000; 2000WO-US13764.

PR 19-MAY-1999; 99US-0134930.

PA (XENC-) XENCOR INC.

PI Dahljat BI;

PT WPI; 2001-025004/03.

DR Non-naturally occurring protein with insulin activity useful for

PT treating type 1 and type 2 diabetes, comprising amino acid

PT substitutions as compared to native human insulin and having enhanced

PT stability -

XX Claim 10; Fig 3C; 95pp; English.

XX The invention relates to novel non-naturally occurring mature human

CC insulin mutants, designated insulin activity (IA) proteins in the

CC specification (AAB48850-B48876), which have altered properties (e.g.,

CC altered oxidative, alkaline or thermal stability, prolonged shelf-life,

CC enhanced ability to oligomerise) compared to wild-type mature human

CC insulin (AAB48848, AAB48849) but which are still capable of binding to

CC an insulin receptor. The insulin mutants of the invention have less than

CC 98% sequence identity to wild-type human insulin, but have a 3D

CC structure which substantially corresponds to that of wild-type human

CC insulin. The invention also relates to recombinant nucleic acids encoding

CC the insulin mutants, expression vectors and host cells comprising mutant

CC insulin-encoding DNA, the recombinant preparation of an insulin mutant,

CC and a pharmaceutical composition comprising an insulin mutant of the

CC invention. The mature human insulin mutants are useful for treating

CC insulin-responsive conditions and disorders of carbohydrate metabolism

CC such as type 1 or type 2 diabetes. The nucleic acids encoding the

CC insulin mutants are used in gene therapy techniques for treating these

CC conditions. The insulin mutants are more stable compared to wild-type

CC mature human insulin, or have an improved ability to form insulin

CC hexamers. Sequences AAB48850-B48863 represent the A and B chains of

CC insulin mutants in which one or more cysteine residue has been

CC replaced, thus preventing the formation of at least one disulphide bond

CC and improving stability and activity.
XX
SQ Sequence 30 AA;

Query Match 94.9%; Score 37; DB 22; Length 30;
Best Local Similarity 87.5%; Pred. No. 0.37;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HLXGSHLV 8
|||
5 hlygshlv 12

RESULT 7

ID AAB48857 standard; peptide; 30 AA.

AC AAB48857;

DT 16-MAR-2001 (first entry)

DE Mutant human insulin (IA protein) cys77d B chain.

XX Human insulin; insulin activity protein; IA protein;

KW diabetes mellitus; type 1; type 2; stability;

KM disulphide bond; cysteine replacement; mutant; mutein.

XX Homo sapiens.

OS Synthetic.

PN WO200069901-A2.

PD 23-NOV-2000.

PF 19-MAY-2000; 2000WO-US13764.

PR 19-MAY-1999; 99US-0134930.

PA (XENC-) XENCOR INC.

PI Dahljat BI;

PT WPI; 2001-025004/03.

DR Non-naturally occurring protein with insulin activity useful for

PT treating type 1 and type 2 diabetes, comprising amino acid

PT substitutions as compared to native human insulin and having enhanced

PT stability -

XX Claim 10; Fig 3D; 95pp; English.

XX The invention relates to novel non-naturally occurring mature human

CC insulin mutants, designated insulin activity (IA) proteins in the

CC specification (AAB48850-B48876), which have altered properties (e.g.,

CC enhanced oxidative, alkaline or thermal stability, prolonged shelf-life,

CC altered ability to oligomerise) compared to wild-type mature human

CC insulin (AAB48848, AAB48849) but which are still capable of binding to

CC an insulin receptor. The insulin mutants of the invention have less than

CC 98% sequence identity to wild-type human insulin, but have a 3D

CC structure which substantially corresponds to that of wild-type human

CC insulin. The invention also relates to recombinant nucleic acids encoding

CC the insulin mutants, expression vectors and host cells comprising mutant

CC insulin-encoding DNA, the recombinant preparation of an insulin mutant,

CC and a pharmaceutical composition comprising an insulin mutant of the

CC invention. The mature human insulin mutants are useful for treating

CC insulin-responsive conditions and disorders of carbohydrate metabolism

CC such as type 1 or type 2 diabetes. The nucleic acids encoding the

CC insulin mutants are used in gene therapy techniques for treating these

CC conditions. The insulin mutants are more stable compared to wild-type
CC mature human insulin, or have an improved ability to form insulin
CC hexamers. Sequences AAB48850-B48863 represent the A and B chains of
CC insulin mutants in which one or more cysteine residue has been
CC replaced, thus preventing the formation of at least one disulphide bond
CC and improving stability and activity.

XX Sequence 30 AA;

Query Match 94.9%; Score 37; DB 22; Length 30;

Best Local Similarity 87.5%; Pred. No. 0.37;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HLXGSHLV 8
| | | | |
Db 5 hlygshlv 12

RESULT 8

AAB48859 ID AAB48859 standard; peptide: 30 AA.

XX AAB48859;

XX 16-MAR-2001 (first entry)

XX Mutant human insulin (IA protein) cyst+ B chain.

XX Human insulin; insulin activity protein; IA protein;

XX diabetes mellitus; type 1; type 2; stability;

XX disulphide bond; cysteine replacement; mutant; mutein.

XX Homo sapiens.

XX Synthetic.

XX WO200069901-A2.

XX 23-NOV-2000.

XX 19-MAY-2000; 2000MO-US13764.

XX 19-MAY-1999; 99US-0134930.

XX (XENC-) XENCOR INC.

XX Dahiyat BI;

XX WPI: 2001-025004/03.

XX Non-naturally occurring protein with insulin activity useful for

XX treating type 1 and type 2 diabetes, comprising amino acid

XX substitutions as compared to native human insulin and having enhanced

XX stability -

XX Claim 10; Fig 3E; 95pp; English.

XX The invention relates to novel non-naturally occurring mature human

XX insulin mutants, designated insulin activity (IA) proteins in the

XX specification (AAB48850-B48876), which have altered properties (e.g.,

XX enhanced oxidative, alkaline or thermal stability, prolonged shelf-life,

XX altered ability to oligomerise) compared to wild-type mature human

XX insulin (AAB48848, AAB48849) but which are still capable of binding to

XX a 98% sequence identity to wild-type human insulin, but have a 3D

XX structure which substantially corresponds to that of wild-type human

XX insulin. The invention also relates to recombinant nucleic acids encoding

XX the insulin mutants, expression vectors and host cells comprising mutant

XX insulin-encoding DNA, the recombinant preparation of an insulin mutant,

XX and a pharmaceutical composition comprising an insulin mutant of the

XX invention. The mature human insulin mutants are useful for treating

XX insulin-responsive conditions and disorders of carbohydrate metabolism

XX such as type 1 or type 2 diabetes. The nucleic acids encoding the

CC insulin mutants are used in gene therapy techniques for treating these
CC conditions. The insulin mutants are more stable compared to wild-type
CC mature human insulin, or have an improved ability to form insulin
CC hexamers. Sequences AAB48850-B48863 represent the A and B chains of
CC insulin mutants in which one or more cysteine residue has been
CC replaced, thus preventing the formation of at least one disulphide bond
CC and improving stability and activity.

XX Sequence 30 AA;

Query Match 94.9%; Score 37; DB 22; Length 30;

Best Local Similarity 87.5%; Pred. No. 0.37;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HLXGSHLV 8
| | | | |
Db 5 hlygshlv 12

RESULT 9

AAB48861 ID AAB48861 standard; peptide: 30 AA.

XX AAB48861;

XX 16-MAR-2001 (first entry)

XX Mutant human insulin (IA protein) helix 24 B chain.

XX Human insulin; insulin activity protein; IA protein;

XX diabetes mellitus; type 1; type 2; stability;

XX disulphide bond; cysteine replacement; mutant; mutein.

XX Homo sapiens.

XX Synthetic.

XX WO200069901-A2.

XX 23-NOV-2000.

XX 19-MAY-2000; 2000MO-US13764.

XX 19-MAY-1999; 99US-0134930.

XX (XENC-) XENCOR INC.

XX Dahiyat BI;

XX WPI: 2001-025004/03.

XX Non-naturally occurring protein with insulin activity useful for

XX treating type 1 and type 2 diabetes, comprising amino acid

XX substitutions as compared to native human insulin and having enhanced

XX stability -

XX Claim 10; Fig 3F; 95pp; English.

XX The invention relates to novel non-naturally occurring mature human

XX insulin mutants, designated insulin activity (IA) proteins in the

XX specification (AAB48850-B48876), which have altered properties (e.g.,

XX enhanced oxidative, alkaline or thermal stability, prolonged shelf-life,

XX altered ability to oligomerise) compared to wild-type mature human

XX insulin (AAB48848, AAB48849) but which are still capable of binding to

XX a 98% sequence identity to wild-type human insulin, but have a 3D

XX structure which substantially corresponds to that of wild-type human

XX insulin. The invention also relates to recombinant nucleic acids encoding

XX the insulin mutants, expression vectors and host cells comprising mutant

XX insulin-encoding DNA, the recombinant preparation of an insulin mutant,

XX and a pharmaceutical composition comprising an insulin mutant of the

XX invention. The mature human insulin mutants are useful for treating

XX insulin-responsive conditions and disorders of carbohydrate metabolism

XX such as type 1 or type 2 diabetes. The nucleic acids encoding the

CC such as type 1 or type 2 diabetes. The nucleic acids encoding the
 CC insulin mutants are used in gene therapy techniques for treating these
 CC mutations. The insulin mutants are more stable compared to wild-type
 CC mature insulin. Sequences AA84830-AA84863 represent the A and B chains of
 CC hexamers. Sequences AA84830-AA84863 represent the A and B chains of
 CC insulin mutants in which one or more cysteine residue has been
 CC replaced, thus preventing the formation of at least one disulphide bond
 CC and improving stability and activity.

XX Sequence 30 AA:

Query Match 94.9% Score 37: DB 22: Length 30:

Best Local Similarity 87.5% Pred. No. 0.37: 1: Indels 0: Gaps 0:

DB 5 HLGSHLV 12

YY 1 HLGSHLV 8

AA82088 standard: peptide: 12 AA.

XX AA82088:

XX 03-APR-1996 (first entry)

XX Pork insulin residues (B) 5-16 T cell epitope.

XX Retro-Inverso modified: T cell epitope; analogue: vaccine: B cell;

XX Immunisation: pork insulin: residues (B) 5-16.

XX Sus scrofa.

XX WO9523166-A1.

XX 31-AUG-1995.

XX 24-FEB-1995: 95WO-A000090.

XX 25-FEB-1994: 94AU-0004119.

XX (DEAK-) DEAKIN RES LTD.

XX Comis A, r P, Tyler MI;

XX WPI: 1995-311503/40.

XX Synthetic peptide T cell epitope analogue (retro-)inverso modified

XX - used conjugated with a B cell epitope in a vaccine tailored to a

XX specific condition, e.g. polio, hepatitis B, etc.

XX Claim 3: Page 32: 51pp; English.

XX A vaccine comprising a retro-inverso modified, native T cell epitope

XX 1, e.g. AA82088, and a B cell epitope (e.g. the malaria

XX immunodominant B-cell epitope (AA82095) and

XX polio/hepatitis, tetanus or hepatitis B proteins etc.) is tailored

XX to a condition of interest (depending on the epitopes used), and

XX used to immunise a host against the specific condition.

XX Sequence 12 AA:

YY 1 HLGSHLV 8

DB 1 HLGSHLV 12

RESULT 11

AA51790 standard: Peptide: 13 AA.

XX AA51790:

XX 18-FEB-1999 (first entry)

XX Synthetic oligonucleotide-labelling peptide.

XX mass spectrometry; detection; identification; diagnosis.

XX Synthetic.

XX Key Location/Qualifiers

XX Modified-site /note= sulfonated cysteine

XX WO926095-A1.

XX 18-JUN-1998.

XX 10-DEC-1997: 97WO-US22639.

XX 16-MAY-1997: 97US-0046710.

XX 10-DEC-1996: 96US-0033037.

XX (GENE-) GENE-TRACE SYSTEMS INC.

XX Becker CH, Montforte JA, Pollart DJ, Shaler TA;

XX WPI: 1998-348547/30.

XX New release tag compounds for detecting target molecule(s) -

XX comprising a reactive group, a release group and a releasable

XX non-volatile mass label detectable by mass spectrometry

XX Example: Page 94: 170pp; English.

XX The sequence is that of a peptide which was used in an

XX oligonucleotide-labelling peptide (e.g. a release

XX tag compound (PTC) these comprise a reactive group, a

XX release group and a non-volatile mass label comprising a

XX synthetic polymer or biopolymer detectable by mass spectrometry.

XX The RTCs can be used as probes for the detection of TMs.

XX Identification of non-coding nucleotide sequences, identification of

XX mutations within a gene or protein sequence, detection of metals, cell,

XX characterisation of antibody-antigen interactions, enzyme-substrate

XX interactions and characterisation of ligand interactions. Multiple

XX applications include multiple pathogen diagnostics, multi-gene genetic

XX polymorphism screening, single nucleotide polymorphism (SNP)

XX genotyping, clone and gene mapping, and gene expression analysis.

XX The RTCs permit the ready detection of releasable mass labels by

XX using a mass spectrometer. The mass labels are prepared by linking

XX that can be used to uniquely identify each desired target.

XX Sequence 13 AA:

YY 1 HLGSHLV 8

DB 5 HLGSHLV 12

Query Match 92.3% Score 36: DB 19: Length 13:

Best Local Similarity 87.5% Pred. No. 0.24: 1: Indels 0: Gaps 0:

DB 5 HLGSHLV 12

YY 1 HLGSHLV 8


```

RESULT 12
AAR51919
ID AAR51919 standard; peptide: 16 AA.
XX
AC AAR51919;
XX
DT 18-JAN-1995 (first entry)
XX
DE Insulin (B1-B16).
XX
KM Insulin: A chain; B chain; conjugate; blood brain barrier; BB3;
  nerve nutrition factor; neuropeptide.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Disulfide-bond 7 /note= "Cys(7) in AAR51918 is disulfide-bonded to
FT Cys(7) in AAR51919"
XX
EP599303-A.
XX
PD 01-JUN-1994.
XX
PF 25-NOV-1993; 93EP-0118961.
XX
PR 27-NOV-1992; 92JP-0318031.
XX
PA (TAKE ) TAKEDA CHEM IND LTD.
XX
PI Fukuta M, Iinuma S, Okada H;
XX
DR WPI; 1994-169627/21.
XX
PT Conjugate for transporting drug across blood brain barrier -
  comprising bioactive peptide or protein and carrier peptide.
XX
PS Claim 6; Fig 3; 10pp; English.
XX
CC A peptide conjugate capable of passing the blood-brain barrier (BBB)
  comprises a bioactive peptide or protein incapable of passing the
  BBB and a carrier peptide which exhibits no bioactivity and is
  capable of passing the BBB. The carrier peptide may be an insulin
  fragment comprising a peptide chain with 14-21 amino acids from the
  N-terminus of insulin chain A and another peptide chain with 16-22
  amino acids from the N-terminus of insulin chain B. The bioactive
  peptide may be a nerve nutrition factor or a neuropeptide.
XX
SQ Sequence 16 AA;

Query Match          92.3%; Score 36; DB 15; Length 16;
Best Local Similarity 87.5%; Pred. No. 0.3;
Matches 7; Conservative 0; Mismatches 1; Indels C; Gaps 0;

OY 1 HLXGSHLV 8
   |||||
DB 5 hlcgshlv 12

RESULT 13
AAR66218
ID AAR66218 standard; peptide: 22 AA.
XX
AC AAR66218;
XX
DT 18-JAN-1995 (first entry)
XX
DE Insulin (B1-B22).
XX
KM Insulin: A chain; B chain; conjugate; blood brain barrier; BBH;
  nerve nutrition factor; neuropeptide.

```

```

XX OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Disulfide-bond 7 /note= "Cys(7) in AAR66217 is disulfide-bonded to
FT Cys(7) in AAR66218"
XX
FT Disulfide-bond 19 /note= "Cys(20) in AAR66217 is disulfide-bonded to
FT Cys(19) in AAR66218"
XX
PN EP599303-A.
XX
PD 01-JUN-1994.
XX
PF 25-NOV-1993; 93EP-0118961.
XX
PR 27-NOV-1992; 92JP-0318031.
XX
PA (TAKE ) TAKEDA CHEM IND LTD.
XX
PI Fukuta M, Iinuma S, Okada H;
XX
DR WPI; 1994-169627/21.
XX
PT Conjugate for transporting drug across blood brain barrier -
  comprising bioactive peptide or protein and carrier peptide.
XX
PS Claim 7; Fig 6; 10pp; English.
XX
CC A peptide conjugate capable of passing the blood-brain barrier (BBB)
  comprises a bioactive peptide or protein incapable of passing the
  BBB and a carrier peptide which exhibits no bioactivity and is
  capable of passing the BBB. The carrier peptide may be an insulin
  fragment comprising a peptide chain with 14-21 amino acids from the
  N-terminus of insulin chain A and another peptide chain with 16-22
  amino acids from the N-terminus of insulin chain B. The bioactive
  peptide may be a nerve nutrition factor or a neuropeptide.
XX
SQ Sequence 22 AA;

Query Match          92.3%; Score 36; DB 15; Length 22;
Best Local Similarity 87.5%; Pred. No. 0.42;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HLXGSHLV 8
   |||||
DB 5 hlcgshlv 12

RESULT 14
AAR75413
ID AAR75413 standard; peptide: 22 AA.
XX
AC AAR75413;
XX
DT 31-JAN-1996 (first entry)
XX
DE Insulin B-chain derivative, desoctapeptide (B23-B30)-insulin.
XX
KM Insulin; diabetes mellitus; detection; therapeutic; auto-antibody.
XX
OS Homo sapiens.
XX
PN US5422339-A.
XX
PD 06-JUN-1995.
XX
PF 19-MAR-1991; 91US-0671455.
XX
PR 19-MAR-1991; 91US-0671455.
  14-MAR-1994; 94US-0212696.

```

XX (JOSL-) JOSLIN DIABETES CENT INC.
 PA Castano L, Eisenbarch GS, Shoelson SE;
 PI WPI; 1995-214691/28.
 XX
 DR
 XX Novel peptide(s) for detecting insulin auto:antibodies and treatment
 PT - have insulin auto:antibody but not insulin receptor binding
 PT capacity
 PS Claim 1: Column 7-8; 10pp; English.
 XX
 CC AAR75412 is the alpha insulin chain and AAR75413 is the insulin B chain
 CC deriv. desoctapeptide(B23-B30)-insulin. These chains form an insulin
 CC peptide which is specifically reactive with human insulin
 CC autoantibodies and non-reactive with human insulin cell surface
 CC receptors. The insulin deriv. is bound to a cytotoxic peptide (either
 CC ricin or diphtheria toxin) and the resulting conjugate is useful for
 CC the detection of human insulin autoantibodies and will destructively
 CC target insulin autoantibodies in serum and autoantibody-producing
 CC B lymphocytes. The conjugate is useful in the prevention of the onset
 CC of Type I diabetes mellitus without triggering the physiological
 CC responses that are a consequence of insulin/receptor binding.
 XX
 SQ Sequence 22 AA;

Query Match 92.3%; Score 36; DB 16; Length 22;
 Best Local Similarity 87.5%; Pred. No. 0.42;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HLXGSHLV 8
 || |||||
 DB 5 hlcsghlv 12

RESULT 15

AAB10773 AAB10773 standard; peptide; 25 AA.
 XX
 AC AAB10773;
 XX
 DT 26-JAN-2001 (first entry)
 DE Insulin beta chain fragment #2.
 XX
 KW Insulin; antidiabetic; treatment; diagnosis; beta-chain.
 XX
 OS Unidentified.
 XX
 PN DE19908041-A1.
 XX
 PD 31-AUG-2000.
 XX
 PF 24-FEB-1999; 99DE-1008041.
 XX
 PR 24-FEB-1999; 99DE-1008041.
 XX
 PA (HOEC/) HOECKER H.
 XX
 PI Brandenburg D, Havenith C;
 XX
 DR WPI; 2000-602964/58.
 XX

PT New insulin dimers useful in the treatment and diagnosis of diabetes
 PT are covalently connected via a dicarboxylic acid -
 XX

PS Claim 1; Page 9; 12pp; German.

CC This invention describes novel insulin analogs which comprise two
 CC insulin molecules, whose N-terminal B-chain amino groups are covalently
 CC connected via a linear bifunctional carboxylic acid of variable length.

CC The products of the invention have antidiabetic activity. The analogs are
 CC useful for the treatment and diagnosis of diabetes. Modification of the
 CC insulin molecule in this way can result in an up to 20-fold increase in
 CC potency compared with native insulin.
 XX
 SQ Sequence 25 AA;

Query Match 92.3%; Score 36; DB 21; Length 25;
 Best Local Similarity 87.5%; Pred. No. 0.48;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HLXGSHLV 8
 || |||||
 DB 5 hlcsghlv 12

Search completed: September 6, 2001, 16:43:29
 Job time: 358 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:49:49 ; Search time 231.42 Seconds
(without alignments)
4.574 Million cell updates/sec

Title: US-09-603-713-9
Perfect score: 40
Sequence: 1 LVNMAEGD 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL_16:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_unclassified:*
13: sp_vertebrate:*
14: sp_virus:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	100.0	469	11	Q91LP9 mus musculus
2	35	87.5	377	2	Q9F6W5 O91P9 chloroflexu
3	34	85.0	219	10	Q9P0Z8 O91P9 nicotiana t
4	34	85.0	854	2	O59556 mycobacteri
5	33	82.5	562	2	P71654 mycobacteri
6	33	82.5	1840	10	Q9SEY4 O91P9 arabidopsi
7	32	80.0	71	2	Q9XCD6 O91P9 mycobacteri
8	32	80.0	292	5	Q27109 O91P9 trichomonas
9	32	80.0	641	2	O52824 O91P9 chlamydia m
10	32	80.0	833	2	O9PKK4 O91P9 chlamydia t
11	32	80.0	836	2	O84192 O91P9 chlamydia t
12	32	80.0	1615	4	O9NYS8 O91P9 homo sapien
13	32	80.0	3016	2	P73590 O91P9 synchocyst
14	31	77.5	55	5	O44360 O91P9 ceratit c
15	31	77.5	90	13	Q92121 O91P9 xiphias gila
16	31	77.5	184	10	O9SJRO O91P9 arabidopsi
17	31	77.5	334	10	O9SP12 O91P9 schizosacch
18	31	77.5	354	3	O60121 O91P9 thermotoga
19	31	77.5	431	2	O9X0X4 O91P9 thermotoga

20	31	77.5	436	2	O91Z76	O91Z76 neisseria m
21	31	77.5	472	5	O17755	O17755 caenorhabd
22	31	77.5	662	10	O81505	O81505 arabidopsi
23	31	77.5	765	1	O59019	O59019 methanococ
24	31	77.5	777	10	O9M912	O9M912 arabidopsi
25	31	77.5	797	2	O9S341	O9S341 photorhabd
26	31	77.5	1420	10	O81016	O81016 arabidopsi
27	31	77.5	181	10	O9M718	O9M718 perseu amer
28	30	75.0	183	10	O9F0Z9	O9F0Z9 nicotiana t
29	30	75.0	188	10	O91N84	O91N84 arabidopsi
30	30	75.0	216	10	O82131	O82131 arabidopsi
31	30	75.0	216	10	O9S4Z3	O9S4Z3 arabidopsi
32	30	75.0	225	14	O36380	O36380 atelaphine
33	30	75.0	314	2	O9ZFM1	O9ZFM1 bacillus st
34	30	75.0	314	10	O65612	O65612 arabidopsi
35	30	75.0	314	10	O9M010	O9M010 arabidopsi
36	30	75.0	335	2	O9PM77	O9PM77 campylobact
37	30	75.0	406	2	O9ZJ14	O9ZJ14 bacillus am
38	30	75.0	474	5	O9V0T4	O9V0T4 drosophila
39	30	75.0	509	14	O9YW96	O9YW96 epiphyas po
40	30	75.0	524	2	O9RYT8	O9RYT8 delinococcus
41	30	75.0	601	10	O9M2H2	O9M2H2 arabidopsi
42	30	75.0	723	2	O9EWC1	O9EWC1 streptomyce
43	30	75.0	724	14	O9O3G5	O9O3G5 turkey asstr
44	30	75.0	750	2	O9KXP6	O9KXP6 streptomyce
45	30	75.0	1194	3	O93962	O93962 glomus vers

ALIGNMENTS

RESULT 1
Q91LP9 PRELIMINARY: PRT: 469 AA.

AC Q91LP9: 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE PRESENTILIN-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SAM P8: TISSUE=HIPPOCAMPUS;
RA Kumar V.B., Vyas K.C., Choudhary V., Franko M., Flood J.F.,
RA Morley J.E.;
RT "Molecular Cloning and Tissue Distribution of Presentilin-1 in
RT Senescence Accelerated Mice (SAM P8) Mice."
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF149111; AAF73153.1; -
DR InterPro: IPR001108; -
DR Pfam: PF01080; Presentilin.1.
DR PRINTS: PR01072; PRESENTILIN.
SQ SEQUENCE 469 AA; 52929 MW; CF92C2A6F398B1DF CRC64;

Query Match 100.0%; Score 40; DB 11; Length 469;
Best Local Similarity 100.0%; Prod. No. 1.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVNMAEGD 8
| | | | | | | | | |
Db 296 LVNMAEGD 303

RESULT 2
Q9F6W5 PRELIMINARY: PRT: 377 AA.
AC Q9F6W5: 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)

DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE HYPOHETHELAL 41.6 KDA PROTEIN (FRAGMENT).
 OS Chloroflexus aurantiacus.
 OC Bacteria; Green non-sulfur bacteria; Chloroflexaceae group;
 CC Chloroflexaceae; Chloroflexus.
 OX NCBI_TaxId=1108;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20433268; PubMed=10976061;
 RA Xiong J., Fischer W.M., Inoue K., Nakahara M., Bauer C.E.;
 RT "Molecular evidence for the early evolution of photosynthesis.";
 RL Science 289:1724-1730(2000).
 DR EMBL; AF288461; AAG15220.1; .
 KW Hypothetical protein.
 FT NON_TER 377
 SQ SEQUENCE 377 AA; 41583 MW; 6F6849081F0E98F8 CRC64;

Query Match 87.5%; Score 35; DB 2; Length 377;
 Best Local Similarity 87.5%; Pred. No. 14;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 LVNMAEGD 8
 Db 210 LVNMAEGD 217

RESULT 3
 ID 09FOZ8 PRELIMINARY; PRT; 219 AA.
 AC 09FOZ8;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DE AVR9/CF-9 RAPIDLY ELICITED PROTEIN 111B.
 GN ACRE11B.
 OS Nicotiana tabacum (Common tobacco).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 CC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
 OX Solanales; Solanaceae; Nicotiana.
 OX NCBI_TaxId=4097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. PETITE HAVANA;
 RA Durand W.E., Rowland O., Piedras P., Hammond-Kosack K.E.,
 RA Jones J.D.G.;
 RT "cDNA expression profiling reveals rapid, resistance gene-dependent,
 RT active oxygen-independent, gene induction during the plant defense
 RT response.";
 RT Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AF211531; AAG43549.1; .
 SQ SEQUENCE 219 AA; 24587 MW; DAF94C81CAC10E54 CRC64;

Query Match 85.0%; Score 34; DB 10; Length 219;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LVNMAEG 7
 Db 185 LVNMAEG 191

RESULT 4
 ID 059556 PRELIMINARY; PRT; 854 AA.
 AC 059556;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DE DNA TOPOISOMERASE (EC 5.99.1.3) (DNA TOPOISOMERASE (ATP-HYDROLYSING))
 DE (DNA TOPOISOMERASE II) (DNA GYRASE) (TYPE II DNA TOPOISOMERASE).
 GN GYR.

OS Mycobacterium smegmatis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 CC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxId=17772;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SN2;
 RX MEDLINE=96118228; PubMed=8574396;
 RA Madhusudan K., Nagaraja V.;
 RT "Mycobacterium smegmatis DNA gyrase: cloning and overexpression in
 RT Escherichia coli.";
 RL Microbiology 141:3029-3037(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SN2;
 RA Valakunja N.;
 RL Submitted (JAN-1995) to the EMBL/Genbank/DBJ databases.
 CC -1- OF DOUBLE-STRANDED DNA.
 CC OF DOUBLE-STRANDED DNA.
 DR EMBL; X84077; CAA58885.1; .
 DR HSP; P09097; IAB4.
 DR InterPro: IPR002205; .
 DR Pfam: PF00521; DNA_topoisolv; 1.
 DR SMART; SM00434; Top4c; 1.
 KW Isomerase.
 SQ SEQUENCE 854 AA; 94924 MW; AEC088EDD51323A1 CRC64;

Query Match 85.0%; Score 34; DB 2; Length 854;
 Best Local Similarity 75.0%; Pred. No. 58;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 LVNMAEGD 8
 Db 803 LVNMAEGD 810

RESULT 5
 ID P71654 PRELIMINARY; PRT; 562 AA.
 AC P71654;
 DT 01-FEB-1997 (TREMBLrel. 02, Created)
 DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
 DE 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 GN RV2797C OR MTCY16B7.46.
 GN RV2797C OR MTCY16B7.46.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 CC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxId=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=96295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
 RA Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 DR EMBL; Z81331; CAB03649.1; .
 DR TubercuList; RV2797C; .
 DR InterPro: IPR000379; .
 DR InterPro: IPR000734; .
 DR PROSITE; PS00120; LIPASE_SER; UNKNOWN_1.
 KW Hypothetical protein.
 SQ SEQUENCE 562 AA; 58791 MW; 5F6B7A6D14F9499F CRC64;

Query Match 82.5%; Score 33; DB 2; Length 562;
 Best Local Similarity 85.7%; Pred. No. 61;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 VNMAEGD 8
 |||:||||
 DB 159 VNMAEGD 165

RESULT 6

Q9SEFY4 PRELIMINARY; PRT; 1840 AA.

AC Q9SEFY4; 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, last sequence update)
 DE 01-MAR-2001 (TREMBlrel. 16, last annotation update)
 DE T22C5.20.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Chao Q., Brooks S., Buehler E., Johnson-Hopson C., Khan S., Kim C.,
 RA Shun P., Altafi H., Bel O., Chin C., Chlou J., Chol E., Conn L.,
 RA Conway A., Gonzales A., Hansen N., Hong B., Koo T., Lam B., Lee J.,
 RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharly N., Nguyen M.,
 RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaverl A.,
 RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
 RA Ecker J.R.;
 RT "Genomic sequence for Arabidopsis thaliana BAC T22C5 from chromosome
 I.";
 RL Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AC012375; AAF24949.1; -;
 DR InterPro: IPR000504; -;
 DR SMART: SM00360; RRM; 1.
 SQ SEQUENCE 1840 AA; 202906 MW; 130ACE4E974B9405 CRC64;

Query Match 82.5%; Score 33; DB 10; Length 1840;
 Best Local Similarity 87.5%; Pred. No. 2.2e+02;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LVNMAEGD 8
 |||:||||
 DB 1713 LVNMAEGD 1720

RESULT 7

Q9XCD6 PRELIMINARY; PRT; 71 AA.

AC Q9XCD6; 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, last annotation update)
 DE HYPOTHETICAL 7.4 KDA PROTEIN.
 OS Mycobacterium smegmatis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1772;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-MC2-155;
 RX MEDLINE-99328972; PubMed-10400584;
 RA Fernandes N.D., Wu Q.-L., Kong D., Hesson R.N.;
 RT "A mycobacterial extracytoplasmic sigma factor involved in survival
 following heat shock and oxidative stress.";
 RL J. Bacteriol. 181:4266-4274 (1999).
 DR EMBL; AF144091; AAD41812.1; -;
 DR HSSP; P10802; IYU.
 DR InterPro: IPR000089; -;

DR Pfam: PF00364; biotin_lipoyl; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 71 AA; 7438 MW; 0E034F0263C7165E CRC64;

Query Match 80.0%; Score 32; DB 2; Length 71;
 Best Local Similarity 85.7%; Pred. No. 10;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 VNMAEGD 8
 |||:||||
 DB 53 VNMAEGD 59

RESULT 8

Q27109 PRELIMINARY; PRT; 292 AA.

AC Q27109; 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, last annotation update)
 DE CYSINE PROTEINASE, PUTATIVE (FRAGMENT).
 GN CP4.
 OS Trichomonas vaginalis.
 OC Eukaryota; Parabasalida; Trichomonadida; Trichomonas.
 OX NCBI_TaxID=5722;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-G3;
 RX MEDLINE-95093621; PubMed-8000542;
 RA Mallinson D.J., Lockwood B.C., Coombs G.H., North M.J.;
 RT "Identification and molecular cloning of four cysteine proteinase
 genes from the pathogenic protozoan Trichomonas vaginalis.";
 RL Microbiology 140:2725-2735(1994).
 DR EMBL; X77221; CA454438.1; -;
 DR HSSP; P07711; ICUL.
 DR MEROPS; C01.082; -;
 DR InterPro: IPR000169; -;
 DR InterPro: IPR000668; -;
 DR Pfam: PF00112; Peptidase_C1; 1.
 DR PRINTS; PR00705; PAPAIN.
 DR PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
 DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
 KW Hydrolase; Thiol protease.
 FT NON TER 1 1
 FT CHAIN 75 292 CYSINE PROTEINASE, PUTATIVE.
 SQ SEQUENCE 292 AA; 32280 MW; 862P69579BD63B45 CRC64;

Query Match 80.0%; Score 32; DB 5; Length 292;
 Best Local Similarity 85.7%; Pred. No. 49;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 VNMAEGD 8
 |||:||||
 DB 187 VNMAEGD 193

RESULT 9

Q52824 PRELIMINARY; PRT; 641 AA.

AC Q52824; 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, last annotation update)
 DE FIXL.
 GN GN
 OS Rhizobium leguminosarum.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 OX NCBI_TaxID=384;
 RN (1)
 RP SEQUENCE FROM N.A.

RC STRAIN-VF39;
 RX MEDLINE-97011572; PubMed-8858582;
 RA Patschkowski T., Schlueter A., Priefer U.;
 RT "Rhizobium leguminosarum bv. viciae contains a second fnr/fixK-like
 gene and an unusual fixL homologue."
 RL Mol. Microbiol. 21:267-280(1996).
 CC -1- SIMILARITY: THE N-TERMINAL REGION IS SIMILAR TO THAT OF OTHER
 CC REGULATORY COMPONENTS OF SENSOR TRANSDUCTION SYSTEMS.
 CC -1- SIMILARITY: TO OTHER PROKARYOTIC SENSOR TRANSDUCTION HISTIDINE
 CC KINASES.
 DR EMBL: Z70305; CAA94319.1; -
 DR HSP; P41789; INTR.
 DR InterPro: IPR000014; -
 DR InterPro: IPR000410; -
 DR InterPro: IPR001789; -
 DR Pfam: PF000072; response_reg; 1.
 DR Pfam: PF000512; signal; 1.
 DR Pfam: PF00989; PAS; 2.
 DR PRINTS: PR00344; BCTRLSENSOR.
 DR SMART: SM00448; REC; 1.
 KW Kinase; Phosphorylation; Sensory transduction; Transferase.
 SQ SEQUENCE 641 AA; 70055 MW; 56714AC1879C9B96 CRC64;

Query Match 80.0%; Score 32; DB 2; Length 641;
 Best Local Similarity 75.0%; Pred. No. 1.2e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 LVNMAEGD 8
 Db 107 LVNLEPGD 114

RESULT 10
 O9PKK4 PRELIMINARY; PRT; 833 AA.
 AC O9PKK4: -
 DT 01-OCT-2000 (TRENBLrel. 15, Created)
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
 DE 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
 DE DNA GYRASE, SUBUNIT A.
 GN TC0461.
 OS Chlamydia muridarum.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83350;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MOPN / NIGS;
 RX MEDLINE-20150255; PubMed-10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.R., Peterson J., Ulterback T., Berry K., Bass S.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
 RT pneumoniae AK39."
 RL Nucleic Acids Res. 28:1397-1406(2000).
 DR EMBL: AE002314; AAF39312.1; -
 DR TIGR: TC0461; -
 DR InterPro: IPR002205; -
 DR Pfam: PF00521; DNA_topoisolv; 1.
 SQ SEQUENCE 833 AA; 93883 MW; D168FE59508D221 CRC64;

Query Match 80.0%; Score 32; DB 2; Length 833;
 Best Local Similarity 75.0%; Pred. No. 1.6e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 LVNMAEGD 8
 Db 795 LVNLEPGD 802

RESULT 11
 ID 084192 PRELIMINARY; PRT; 836 AA.
 AC 084192;
 DT 01-NOV-1998 (TRENBLrel. 08, Created)
 DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)
 DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
 DE DNA GYRASE SUBUNIT A.
 GN GYRA_1.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=813;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-D/W-3/CX;
 RX MEDLINE-99000809; PubMed-9784136;
 RA Stephens R.S., Kaiman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
 RA Mitchell W.P., Olinger L., Tatunov R.L., Zhao Q., Koonin E.V.,
 RA Davis R.W.;
 RT "Genome sequence of an obligate intracellular pathogen of humans:
 RT Chlamydia trachomatis."
 RL Science 282:754-759(1998).
 RN [2]
 RP SEQUENCE OF 1-179 FROM N.A.
 RC STRAIN-L2;
 RA Desus-Babus S., Bebear C.M., Charron A., Bebear C., de Barbeyrac B.;
 RT "Sequencing of gyrase and topoisomerase IV QRDS of Chlamydia
 RT trachomatis and characterization of quinolone-resistant mutants
 RT obtained in vitro."
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF001292; AAC67781.1; -
 DR EMBL: AF044267; AAC33552.1; -
 DR HSP; P09097; IAB4.
 DR InterPro: IPR002205; -
 DR Pfam: PF00521; DNA_topoisolv; 1.
 DR SMART: SM00434; TOP4C; 1.
 SQ SEQUENCE 836 AA; 94232 MW; AA750FA55138B5FC CRC64;

Query Match 80.0%; Score 32; DB 2; Length 836;
 Best Local Similarity 75.0%; Pred. No. 1.6e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 LVNMAEGD 8
 Db 795 LVNLEPGD 802

RESULT 12
 ID 09NYS8 PRELIMINARY; PRT; 1615 AA.
 AC 09NYS8;
 DT 01-OCT-2000 (TRENBLrel. 15, Created)
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
 DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
 DE CLASS III MYOSIN.
 GN MYO3A.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Dose A.C., Burnside B.;
 RT "Cloning and chromosomal localization of a human class III myosin."
 RL Genomics 0:0-0(2000).
 DR EMBL: AF229172; AAF70861.1; -
 DR InterPro: IPR000048; -
 DR InterPro: IPR000719; -
 DR InterPro: IPR001609; -
 DR Pfam: PF00063; myosin_head; 1.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF00612; IO; 3.

DR PRINTS; PR00193; MYOSINHEAVY.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
 DR SMART; SM00015; IQ; 1.
 KM ATP-binding; Transferase.
 SQ SEQUENCE 1615 AA; 185967 MW; CDAEB6CPC379E27 CRC64;

Query Match 80.0%; Score 32; DB 4; Length 1615;
 Best Local Similarity 62.5%; Pred. No. 3.2e+02;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 LVNMAECD 8
 1:1:1:1:1
 DB 899 LVNMAECD 906

RESULT 13
 P73590 PRELIMINARY; PRT; 3016 AA.
 AC P73590;
 DT 01-FEB-1997 (TREMBlrel. 02, Created)
 DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
 DE 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE HYPOTHETICAL 311.5 KDA PROTEIN.
 GN SLR1403.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 OX NCBI_Taxid=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-97061201; PubMed-8905231;
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
 RA Miyajima N., Hirosewa M., Suglura M., Sasamoto S., Kinura T.,
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
 RA Shimp S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
 RA Tabata S.;
 RT *Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions.";
 RL DNA Res. 3:109-116(1996).
 DR EMBL; D90907; BA017634.1; -;
 DR InterPro; IPR000413; -;
 DR InterPro; IPR001343; -;
 DR InterPro; IPR003355; -;
 DR InterPro; IPR003644; -;
 DR Pfam; PF01839; FG-GAP; 8.
 DR PRINTS; PR00313; CABINDGPT.
 DR PRINTS; PR01185; INTEGRINA.
 DR SMART; SM00237; Calx_beta; 1.
 KM Hypothetical protein.
 SQ SEQUENCE 3016 AA; 311507 MW; B34DB3B005D0717A CRC64;

Query Match 80.0%; Score 32; DB 2; Length 3016;
 Best Local Similarity 75.0%; Pred. No. 6.4e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 LVNMAECD 8
 1:1:1:1:1
 DB 1659 LVNMAECD 1666

RESULT 14
 O44360 PRELIMINARY; PRT; 55 AA.
 AC O44360;
 DT 01-JUN-1998 (TREMBlrel. 06, Created)
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
 DE CYTOCHROME P450 MONOOXYGENASE (DU154CC) (FRAGMENT).
 GN CYP6A11.
 OS Ceraulitis capitata (Mediterranean fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Tephritoidea; Tephritidae; Ceraulitis.
 OX NCBI_Taxid=7213;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Danielson P.B., Foster J.L.M., Cooper S.K., Fogelman J.C.;
 RL Submitted (Oct-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) -> ROH +
 CC OXIDIZED FLAVOPROTEIN + H(2)O.
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY
 CC SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL; AF028002; AAB94119.1; -;
 KM Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
 KM Microsome; Endoplasmic reticulum.
 FT NON_TER 1
 SQ SEQUENCE 55 AA; 6332 MW; D182E26FF3850513 CRC64;

Query Match 77.5%; Score 31; DB 5; Length 55;
 Best Local Similarity 85.7%; Pred. No. 13;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 VVMAECD 8
 1:1:1:1:1
 DB 39 VVMAECD 45

RESULT 15
 O92121 PRELIMINARY; PRT; 90 AA.
 ID O92121;
 AC O92121;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE SARCOMPLASMIC RETICULUM CA2+ ATPASE, SLOW ISOFORM (FRAGMENT).
 GN SERCA2.
 OS Xiphias gladius (Swordfish).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Scombroidei;
 OC Xiphiidae; Xiphiidae; Xiphas.
 OX NCBI_Taxid=8245;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-96331512; PubMed-8760229;
 RA Tullis A., Block B.A.;
 RT "Expression of sarcoplasmic reticulum Ca(2+)-ATPase isoforms in marlin
 RT and swordfish muscle and heater cells.";
 RL Am. J. Physiol. 271:R262-R275(1996).
 DR EMBL; U58325; AAB17079.1; -;
 DR HSP; P11719; TEU1.
 DR InterPro; IPR001757; -;
 DR PROSITE; PS00154; ATPASE_E1_E2; UNKNOWN.1.
 FT NON_TER 1
 SQ SEQUENCE 90 AA; 9709 MW; EA221B74786D10D7 CRC64;

Query Match 77.5%; Score 31; DB 13; Length 90;
 Best Local Similarity 75.0%; Pred. No. 22;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 LVNMAECD 8
 1:1:1:1:1
 DB 20 VVMAECD 27

Search completed: September 6, 2001, 16:49:50
 Job time: 734 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:51:08 ; Search time 72.75 Seconds
(without alignments)
3.767 Million cell updates/sec

Title: US-09-603-713-9

Perfect score: 40

Sequence: 1 LVNMAEGD 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486-residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	100.0	467	1 PSN1_HUMAN	P49768 homo sapien
2	40	100.0	467	1 PSN1_MICMU	P79602 microcebus
3	40	100.0	467	1 PSN1_MOUSE	P49769 mus musculu
4	40	100.0	468	1 PSN1_RAT	P97887 ratus norv
5	40	100.0	478	1 PSN1_BOVIN	O9X197 bos taurus
6	37	92.5	433	1 PSN1_XENLA	O12972 xenopus lae
7	34	85.0	842	1 GYRA_MYCSM	P48354 mycobacteri
8	32	80.0	674	1 GSPD_VIBCH	P45779 vibrio chol
9	31	77.5	317	1 ROH3_RAT	P55006 ratus norv
10	31	77.5	404	1 SAHL_THEMA	O51933 thermotoga
11	30	75.0	392	1 MFT1_YEAST	P33441 saccharomyc
12	30	75.0	406	1 ARG1_BACSU	P36843 b atyline
13	30	75.0	679	1 GR75_CRIGR	O35501 cricetulus
14	30	75.0	679	1 GR75_HUMAN	P38646 homo sapien
15	30	75.0	679	1 GR75_MOUSE	P38647 mus musculu
16	30	75.0	679	1 GR75_RAT	P48721 ratus norv
17	30	75.0	723	1 PABS_STRGR	P32483 streptomyce
18	30	75.0	838	1 GYRA_MYCTU	O07702 mycobacteri
19	30	75.0	859	1 LUXO_VIBHA	P54302 vibrio harv
20	30	75.0	1273	1 GYRA_MYCLE	O57532 mycobacteri
21	30	75.0	1902	1 SMFL_HUMAN	O14497 homo sapien
22	30	75.0	2386	1 RAD3_SCHPO	O20095 schizosacch
23	29	72.5	233	1 GP42_RAT	P23505 ratus norv
24	29	72.5	268	1 YBGF_PSEPU	P43037 pseudomonas
25	29	72.5	273	1 RLA_AERPE	O9YFM1 aeropyrum p
26	29	72.5	278	1 D7_XENLA	P13307 xenopus lae
27	29	72.5	281	1 YJCZ_ECOLI	P39667 escherichia
28	29	72.5	319	1 PP12_ACECL	P48481 acetabulari
29	29	72.5	398	1 Y666_METJA	O5880 methanococc
30	29	72.5	457	1 SYS_SUTSO	O33780 sulfolobus
31	29	72.5	469	1 BGL1_BACSU	P40740 bacillus su
32	29	72.5	510	1 MURF_MYCTU	O06220 mycobacteri
33	29	72.5	743	1 CUI2_CAEEL	O17390 caenorhabdi

34	29	72.5	829	1 IP2_HAEIN	P44323 haemophilus
35	28	70.0	98	1 FER1_SYN7	P06517 synchococc
36	28	70.0	138	1 YRRK_BACSU	O34634 bacillus su
37	28	70.0	148	1 YB44_SCHPO	O14350 schizosacch
38	28	70.0	224	1 GT26_DROME	O9V93 drosophila
39	28	70.0	265	1 MUR1_VIBCH	O9X17 vibrio chol
40	28	70.0	277	1 XAPA_ECOLI	P43563 escherichia
41	28	70.0	286	1 RIP1_MOMCH	P16094 momodica c
42	28	70.0	346	1 YJ95_YEAST	P47181 saccharomyc
43	28	70.0	357	1 PD2R_MOUSE	P70263 mus musculu
44	28	70.0	366	1 APN1_SCHPO	P50525 schizosacch
45	28	70.0	391	1 PURT_ECOLI	P33221 escherichia

ALIGNMENTS

RESULT 1
ID PSN1_HUMAN STANDARD; PRT; 467 AA.
AC P49768; Q14762; Q15719; Q15720;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PRESENILIN 1 (PS-1) (S182 PROTEIN).
GN PSEN1 OR PSNLI OR AD3 OR PSI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANTS AD (ISOFORMS I-467 AND I-463).
RX TISSUE=Brain;
MEDLINE=9519502; PubMed=7596406;
RA Sherrington R., Rogaev E.I., Liang Y., Rogaeva E.A., Levesque G., Ikeda M., Chui H., Lin C., Li G., Holman K., Tsuda T., Mar L., Foncin J.-F., Bruni A.C., Montesi M.P., Sorbi S., Rahero I., Pinessi L., Nee L., Chumakov I., Pollen D., Brookes A., Sausseau P., Polinsky R.J., Masco W., da Silva H.A.R., Haines J.L., Perlick Vance M.A., Tanzi R.E., Roses A.D., Fraser P.E., Rommens J.M., St George-Hyslop P.H.;
RA "Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease."
RT Nature 375:754-760(1995).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS I-463 AND I-374).
RX TISSUE=Blood, and Brain;
MEDLINE=96193901; PubMed=8641442;
RA Sahara N., Yahagi Y.-I., Takagi H., Kondo T., Okochi M., Usami M., Shirasawa T., Mori H.;
RT "Identification and characterization of presenilin I-467, I-463 and I-374."
RL FEBS Lett. 381:7-11(1996).
RN [3]
RP SEQUENCE FROM N.A.
RA Rowen L., Madan A., Qin S., Abassi N., Dors M., Ratcliffe A., Madan A., Dickhoff R., Shaffer T., James R., Laaky S., Hood L.;
RT "Complete sequence of the gene for presenilin 1."
RN Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE OF 1-113 FROM N.A.
RA Tsujimura A., Hashimoto-Gotoh T.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SUBCELLULAR LOCATION, AND TISSUE SPECIFICITY.
RX MEDLINE=96160372; PubMed=8574969;
RA Kovacs D.E., Fausett H.J., Page K.J., Kim T.-W., Molr R.D., Merriam D.E., Hollister R.D., Hallmark O.G., Mancini R., Felsenstein K.M., Hyman B.T., Tanzi R.E., Masco W.;
RT "Alzheimer-associated presenilins 1 and 2: neuronal expression in brain and localization to intracellular membranes in mammalian cells."
RL Nat. Med. 2:224-229(1996).

[6]
 RP REVIEW ON VARIANTS.
 RX MEDLINE-97029239; PubMed-8875251;
 RA Cruts M., Hendriks L., Van Broeckhoven C.;
 RT "The presenilin genes: a new gene family involved in Alzheimer disease
 pathology.";
 RL Hum. Mol. Genet. 5:1449-1455(1996).
 [7]
 RP REVIEW ON VARIANTS.
 RX MEDLINE-98180715; PubMed-9521418;
 RA Cruts M., van Broeckhoven C.;
 RT "Presenilin mutations in Alzheimer's disease.";
 RL Hum. Mutat. 11:183-190(1998).
 [8]
 RP VARIANTS AD THR-143 AND ALA-384.
 RX MEDLINE-96177673; PubMed-8634711;
 RA Cruts M., Backovens H., Wang S.-Y., van Gassen G., Theuns J.,
 Breyland M., Datsen N., Weissendach J., den Dunnen J.T., Martin J.-J.,
 Hendriks L., Van Broeckhoven C.;
 RT "Molecular genetic analysis of familial early-onset Alzheimer's
 disease linked to chromosome 14q24.3.";
 RL Hum. Mol. Genet. 4:2363-2372(1995).
 [9]
 RP VARIANTS AD L-82; H-115; T-139; R-163; T-231; L-264; V-392 AND Y-410.
 RX MEDLINE-96177674; PubMed-8634712;
 RA Campion D., Flaman J.-M., Brice A., Hannequin D., Dubois B.,
 Martin C., Moreau V., Charbonnier F., Didierjean O., Tardieu S.,
 Penet C., Puel M., Pasquier F., Le Dore F., Bellis G., Calenda A.,
 Hellig R., Martinez M., Mallet J., Bellis M., Clerget-Daroux F.,
 Aïd J., Frebourg T.;
 RT "Mutations of the presenilin 1 gene in families with early-onset
 Alzheimer's disease.";
 RL Hum. Mol. Genet. 4:2373-2377(1995).
 [10]
 RP VARIANTS AD VAL-260; VAL-285 AND VAL-392.
 RX MEDLINE-95379711; PubMed-765136;
 RA Rogaeve E.I., Sherrington R., Rogaeve E.A., Levesque G., Ikeda M.,
 Liang Y., Chi H., Lin C., Holman K., Tsuda T., Mar L., Sorbi S.,
 Nacmas B., Placentini S., Amaducci J.M., Chunaikov I., Cohen D.,
 Lanfelle L., Fraser P.E., Kommen J.M., St George-Hyslop P.H.;
 RT "Familial Alzheimer's disease in kindreds with missense mutations in
 a gene on chromosome 1 related to the Alzheimer's disease type 3
 gene.";
 RL Nature 376:775-778(1995).
 [11]
 RP VARIANTS AD V-139; V-146; T-163; T-267; A-280 AND G-280.
 RX MEDLINE-96024664; PubMed-7550356;
 RA Claret R.F., Hutton M., Fuldner R.A., Froelich S., Karan E.,
 Talbot C., Crook R., London C., Prihar G., He C., Korenblatt K.,
 Martinez A., Wragg M., Busfield F., Behrens M.I., Myers A., Norton J.,
 Morris J., Mehta N., Pearson C., Lincoln S., Baker M., Duff K.,
 Zehr C., Perez-Tur J., Houlden H., Ruiz A., Ossa J., Lopera F.,
 Arcos M., Madrigal L., Collinge J., Humphreys C., Asworth T.,
 Sarter S., Fox N., Harvey R., Kennedy A., Roques P., Cline R.T.,
 Phillips C.A., Ventur R., Villanen M., Winblad B., Lilius L.,
 Johnson J., Compton R., Forsell L., Axelman K., Kallio K., Haltia M.,
 Poytonen M., Dickson D., Mann D., Neary D., Snowden J., Lantos P.,
 Lanfelle L., Rossor M., Roberts G.W., Adams M.D., Hardy J., Goate A.;
 RT "The structure of the presenilin 1 (S182) gene and identification of
 six novel mutations in early onset AD families.";
 RL Nat. Genet. 11:219-222(1995).
 [12]
 RP VARIANTS AD PHE-96; ARG-163 AND THR-213.
 RX MEDLINE-96310408; PubMed-873303;
 RA Kamino K., Sato S., Sakaki Y., Yoshiwa A., Nishiwaki Y., Takeda H.,
 Tanabe H., Nishimura T., Ii K., St George-Hyslop P.H., Miki T.,
 Ogihara T.;
 RT "Three different mutations of presenilin 1 gene in early-onset
 Alzheimer's disease families.";
 RL Neurosci. Lett. 208:195-198(1996).
 [13]
 RP VARIANT AD ASP-135.

RX MEDLINE-97369208; PubMed-9225696;
 RA Crook R., Ellis R., Shanks M., Thal L.J., Perez-Tur J., Baker M.,
 Hutton M., Haltia T., Hardy J., Galasko D.;
 RT "Early-onset Alzheimer's disease with a presenilin-1 mutation at the
 site corresponding to the Voigt German presenilin-2 mutation.";
 RL Ann. Neurol. 42:124-128(1997).
 [14]
 RP VARIANT AD ALA-280.
 RX MEDLINE-97442268; PubMed-9298817;
 RA London C.L., Martinez A., Behrens I.M., Kosik K.S., Madrigal L.,
 Norton J., Neuman R., Myers A., Busfield F., Wragg M., Arcos M.,
 Arango Viana J.C., Ossa J., Ruiz A., Goate A.M., Lopera F.;
 RT "E280A PS-1 mutation causes Alzheimer's disease but age of onset is
 not modified by APOE alleles.";
 RL Hum. Mutat. 10:186-195(1997).
 [15]
 RP VARIANTS AD THR-233 AND THR-278.
 RX MEDLINE-97316242; PubMed-9172170;
 RA Kwok J.B.J., Tadel K., Hallupp M., Fisher C., Brooks W.S., Broe G.A.,
 Hardy J., Fulham M.J., Nicholson G.A., Stell R.,
 St George-Hyslop P.H., Fraser P.E., Kakulas B., Clarette R.,
 Relkin N., Gandy S.E., Schofield P.R., Martins R.N.;
 RT "Two novel (M233T and R278T) presenilin-1 mutations in early-onset
 Alzheimer's disease pedigrees and preliminary evidence for
 association of presenilin-1 mutations with a novel phenotype.";
 RL NeuroReport 8:1537-1542(1997).
 [16]
 RP VARIANT AD PRO-171.
 RX MEDLINE-99050262; PubMed-9833068;
 RA Ramirez-Duenas M.G., Rogaeve E.A., Leal C.A., Lin C.,
 Ramirez-Castillas G.A., Hernandez-Romo J.A., St George-Hyslop P.H.,
 Cantu J.M.;
 RT "A novel Leu171Pro mutation in presenilin-1 gene in a Mexican family
 with early onset Alzheimer disease.";
 RL Ann. Genet. 41:149-153(1998).
 [17]
 RP VARIANT GLY-318.
 RX MEDLINE-99066768; PubMed-9851443;
 RA Mattila K.M., Forsell C., Pirttila T., Rime J.O., Lehtimäki T.,
 Roytta M., Lilius L., Berola A., St George-Hyslop P.H., Frey H.,
 Lanfelle L.;
 RT "The Glu318Gly mutation of the presenilin-1 gene does not necessarily
 cause Alzheimer's disease.";
 RL Ann. Neurol. 44:965-967(1998).
 [18]
 RP VARIANT GLY-318.
 RX MEDLINE-99066775; PubMed-9851450;
 RA Aldudo J., Bullido M.J., Frank A., Valdivieso F.;
 RT "Missense mutation E318G of the presenilin-1 gene appears to be a
 nonpathogenic polymorphism.";
 RL Ann. Neurol. 44:985-986(1998).
 [19]
 RP VARIANTS AD VAL-79; CYS-115; VAL-231, AND VARIANT GLY-318.
 RX MEDLINE-98046005; PubMed-9384602;
 RA Cruts M., van Duin C.M., Backovens H., van den Broeck M.,
 Weinert A., Seneels S., Sherrington R., Hutton M., Hardy J.,
 St George-Hyslop P.H., Hofman A., van Broeckhoven C.;
 RT "Estimation of the genetic contribution of presenilin-1 and -2
 mutations in a population-based study of presenile Alzheimer
 disease.";
 RL Hum. Mol. Genet. 7:43-51(1998).
 [20]
 RP VARIANTS AD D-120; R-163; V-209; V-260; L-264; Y-410 AND P-426.
 RX MEDLINE-98180720; PubMed-9521423;
 RA Poorakaj P., Sharma V., Anderson L., Nemens E., Alonso M.E., Orr H.,
 White J., Heston L., Bird T.D., Schellenberg G.D.;
 RT "Missense mutations in the chromosome 14 familial Alzheimer's disease
 presenilin 1 gene.";
 RL Hum. Mutat. 11:216-221(1998).
 [21]
 RP VARIANT AD GLU-378.
 RX MEDLINE-99212125; PubMed-10200054;
 RA Besancon R., Lorenzi A., Cruts M., Radawiec S., Sturtz F.,

Query Match 100.0%; Score 40; DB 1; Length 467;
 Best Local Similarity 100.0%; Pred. No. 0.21;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LVNMAEGD 8
 |||||
 DB 295 LVNMAEGD 302

RESULT 2
 PSN1_MICMU STANDARD: PRT: 467 AA.
 AC P79802;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE PRESENTILIN 1 (PS-1).
 GN PSEN1 OR PSN1 OR PSL.
 OS Microcebus murinus (lesser mouse lemur).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Strepsirrhini; Cheirogaleidae;
 OC Microcebus.
 OX NCBI_TaxID=30608;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Brain;
 RX MEDLINE-97079199; PubMed-8920931;
 RA Calenda A., Westre-Frances N., Czech C., Pradler L., Bons N.,
 RA Bellis M.;
 RT "Molecular cloning, sequencing, and brain expression of the
 RT Presentilin 1 gene in Microcebus murinus";
 RL Biochem. Biophys. Res. Commun. 228:430-439(1996).
 CC -1- FUNCTION: MAY PLAY A ROLE IN INTRACELLULAR SIGNALING AND GENE
 CC EXPRESSION OR IN LINKING CHROMATIN TO THE NUCLEAR MEMBRANE.
 CC MAY FUNCTION IN THE CYTOPLASMIC PARTITIONING OF PROTEINS.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. GOLGI AND
 CC ENDOPLASMIC RETICULUM (BY SIMILARITY).
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: I-467 (SHOWN HERE) AND I-
 CC 463; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: FOUND PREDOMINANTLY IN NEURONS OF THE
 CC DIFFERENT CORTICAL LAYERS AND HIPPOCAMPUS BUT ALSO IN SUBCORTICAL
 CC STRUCTURES.
 CC -1- SIMILARITY: BELONGS TO THE PRESENTILIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: 271333; CA95930.1; -
 DR MEROPS: A22.001; -
 DR InterPro: IPR001108; -
 DR InterPro: IPR002031; -
 DR Pfam: PF01080; Presentilin-1.
 DR PRINTS: PRO1072; PRESENTILIN.
 DR Transmembrane: Glycoprotein; Endoplasmic reticulum; Golgi stack;
 KW Alternative splicing.
 CC FT TRANSMEM 83 103 POTENTIAL.
 CC FT TRANSMEM 133 153 POTENTIAL.
 CC FT TRANSMEM 161 181 POTENTIAL.
 CC FT TRANSMEM 191 211 POTENTIAL.
 CC FT TRANSMEM 221 241 POTENTIAL.
 CC FT TRANSMEM 244 264 POTENTIAL.
 CC FT TRANSMEM 281 301 POTENTIAL.
 CC FT TRANSMEM 408 428 POTENTIAL.
 CC FT TRANSMEM 433 453 POTENTIAL.
 CC FT CARBOHYD 279 279 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 405 405 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT VARSPIC 26 29 MISSING (IN ISOFORM I-463).
 SQ SEQUENCE 467 AA; 52384 MB; D966FF2CA/F2375C CRC64;

Query Match 100.0%; Score 40; DB 1; Length 467;
 Best Local Similarity 100.0%; Pred. No. 0.21;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LVNMAEGD 8
 |||||
 DB 295 LVNMAEGD 302

RESULT 3
 PSN1_MOUSE STANDARD: PRT: 467 AA.
 ID PSN1_MOUSE
 AC P49769;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE PRESENTILIN 1 (PS-1) (S182 PROTEIN).
 GN PSEN1 OR PSN1 OR AD3H.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Brain;
 RX MEDLINE-95319502; PubMed-7596406;
 RA Sherrington R., Rogaev E.I., Liang Y., Rogaeva E.A., Levesque G.,
 RA Ikeda M., Chi H., Lin C., Li G., Holman K., Tsuda T., Mar L.,
 RA Foncin J.-F., Bruni A.C., Montesi M.P., Sorbi S., Rahner I.,
 RA Pinessi L., Nee L., Chumakov I., Pollen D., Brookes A.,
 RA Saneau P., Polinsky R.J., Masco W., da Silva H.A.R., Haines J.L.,
 RA Perleak Vance M.A., Tanzi R.E., Roses A.D., Fraser P.E.,
 RA Rommens J.M., St George-Hyslop P.H.;
 RT "Cloning of a gene bearing missense mutations in early-onset familial
 RT Alzheimer's disease";
 RL Nature 375:754-760(1995).
 CC [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-129/SVJ;
 RX MEDLINE-97442406; PubMed-9295283;
 RA Mitsuda N., Roses A.D., Vittek M.P.;
 RT "Transcriptional regulation of the mouse presentilin-1 gene";
 RL J. Biol. Chem. 272:23489-23497(1997).
 CC -1- FUNCTION: MAY PLAY A ROLE IN INTRACELLULAR SIGNALING AND GENE
 CC EXPRESSION OR IN LINKING CHROMATIN TO THE NUCLEAR MEMBRANE.
 CC MAY FUNCTION IN THE CYTOPLASMIC PARTITIONING OF PROTEINS.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. GOLGI AND
 CC ENDOPLASMIC RETICULUM (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE PRESENTILIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: LA2177; AAC42094.1; -
 DR MEROPS: AF007560; AAB72049.1; -
 DR MEROPS: A22.001; -
 DR MGD: MGI:1202717; Psen1.
 DR InterPro: IPR001108; -
 DR InterPro: IPR002031; -
 DR Pfam: PF01080; Presentilin-1.
 DR PRINTS: PRO1072; PRESENTILIN.
 DR PRINTS: PRO1073; PRESENTILIN.
 KW Transmembrane; Glycoprotein; Endoplasmic reticulum; Golgi stack.
 CC FT TRANSMEM 83 103 POTENTIAL.

```

FT TRANSMEM 133 153 POTENTIAL.
FT TRANSMEM 161 181 POTENTIAL.
FT TRANSMEM 195 215 POTENTIAL.
FT TRANSMEM 221 241 POTENTIAL.
FT TRANSMEM 244 264 POTENTIAL.
FT TRANSMEM 281 301 POTENTIAL.
FT TRANSMEM 408 428 POTENTIAL.
FT TRANSMEM 433 453 POTENTIAL.
FT CARBOHYD 30 30 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 279 279 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 405 405 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 467 AA; 52639 MM; D07215B4BAD2D549 CRC64;

Query Match 100.0%; Score 40; DB 1; Length 467;
Best Local Similarity 100.0%; Pred. No. 0.21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LVNMAEGD 8
Db 295 LVNMAEGD 302

RESULT 4
PSNL_RAT STANDARD: PRT: 468 AA.
ID PSNL_RAT
AC P97887; P97529;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PRESENTILIN 1 (PS-1) (S182 PROTEIN).
GN PSEN1 OR PSNL1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
ON NCBI_TaxID=10116;
RX MEDLINE=97199371; PubMed=9047347;
RA Taniguchi T., Hashimoto T., Taniguchi R., Shimada K., Kawamata T.,
RA Yasuda M., Nakai M., Terashima A., Koizumi T., Maeda K., Tanaka C.;
RT "Cloning of the cDNA encoding rat presentilin-1."
RL Gene 186:73-75(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-WISTAR; TISSUE-Brain;
RA MEDLINE=96255262; PubMed=8710164;
RA Takahashi H., Murayama M., Takashima A., Mercken M., Nakazato Y.,
RA Noguchi K., Imahori K.;
RT "Molecular cloning and expression of the rat homologue of
RT presentilin-1."
RL Neurosci. Lett. 206:113-116(1996).
CC -1- FUNCTION: MAY PLAY A ROLE IN INTRACELLULAR SIGNALING AND GENE
CC EXPRESSION OR IN LINKING CHROMATIN TO THE NUCLEAR MEMBRANE.
CC MAY FUNCTION IN THE CYTOPLASMIC PARTITIONING OF PROTEINS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. GOLGI AND
CC ENDOPLASMIC RETICULUM (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE PRESENTILIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; D82578; BA011575.1; -
CC DR EMBL; D82578; BA011575.1; -
CC DR MEROPS; A22.001; -
CC DR InterPro; IPR001108; -
CC DR InterPro; IPR002031; -

```

```

DR Pfam; PF01080; Presentilin; 1.
DR PRINTS; PR01072; PRESENTILIN.
DR PRINTS; PR01073; PRESENTILIN.
KW Transmembrane; Glycoprotein; Endoplasmic reticulum; Golgi stack.
FT TRANSMEM 83 103 POTENTIAL.
FT TRANSMEM 133 153 POTENTIAL.
FT TRANSMEM 161 181 POTENTIAL.
FT TRANSMEM 195 215 POTENTIAL.
FT TRANSMEM 221 241 POTENTIAL.
FT TRANSMEM 244 264 POTENTIAL.
FT TRANSMEM 281 301 POTENTIAL.
FT TRANSMEM 409 429 POTENTIAL.
FT TRANSMEM 434 454 POTENTIAL.
FT CARBOHYD 57 57 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 279 279 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 406 406 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 234 234 A -> S (IN REF. 2).
FT CONFLICT 381 381 K -> R (IN REF. 2).
SQ SEQUENCE 468 AA; 52790 MM; 17CB791E88A16FC0 CRC64;

Query Match 100.0%; Score 40; DB 1; Length 468;
Best Local Similarity 100.0%; Pred. No. 0.21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LVNMAEGD 8
Db 295 LVNMAEGD 302

RESULT 5
PSNL_BOVIN STANDARD: PRT: 478 AA.
ID PSNL_BOVIN
AC O9XT97;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PRESENTILIN 1 (PS-1).
GN PSEN1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
ON NCBI_TaxID=9913;
RX MEDLINE=9913;
RA Sahara N., Shirasawa T., Mori H.;
RT "Molecular cloning of bovine presentilin 1 gene."
RT Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY PLAY A ROLE IN INTRACELLULAR SIGNALING AND GENE
CC EXPRESSION OR IN LINKING CHROMATIN TO THE NUCLEAR MEMBRANE. MAY
CC FUNCTION IN THE CYTOPLASMIC PARTITIONING OF PROTEINS (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. GOLGI AND
CC ENDOPLASMIC RETICULUM (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE PRESENTILIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AF038936; AAD39023.1; -
CC DR EMBL; AF038936; AAD39023.1; -
CC DR InterPro; IPR001108; -
CC DR InterPro; IPR002031; -
CC DR Pfam; PF01080; Presentilin; 1.
CC DR PRINTS; PR01072; PRESENTILIN.
CC DR PRINTS; PR01073; PRESENTILIN.
KW Transmembrane; Glycoprotein; Endoplasmic reticulum; Golgi stack.

```

FT TRANSMEM 83 103 POTENTIAL.
 FT TRANSMEM 134 154 POTENTIAL.
 FT TRANSMEM 162 182 POTENTIAL.
 FT TRANSMEM 196 216 POTENTIAL.
 FT TRANSMEM 222 242 POTENTIAL.
 FT TRANSMEM 245 265 POTENTIAL.
 FT TRANSMEM 282 302 POTENTIAL.
 FT TRANSMEM 392 412 POTENTIAL.
 FT TRANSMEM 419 439 POTENTIAL.
 FT TRANSMEM 444 464 POTENTIAL.
 FT CARBOHYD 30 30 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 280 280 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 416 416 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 478 AA: 53653 MW: 59E3FC0A1010D906 CRC64:

Query Match Best Local Similarity 100.0%; Score 40; DB 1; Length 478;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LVNMAEGD 8
 |||||
 Db 296 LVNMAEGD 303

RESULT 6
 ID PSN1_XENLA STANDARD: PRT: 433 AA.
 AC 012976;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE PRESENTILIN ALPHA.
 GN PS-ALPHA.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=97223465; PubMed=9070286;
 RA Tsujimura A., Yasojima K., Hashimoto-Gotoh T.;
 RT "Cloning of xenopus presentilin-alpha and -beta cDNAs and their
 differential expression in oogenesis and embryogenesis.";
 RL Biochem. Biophys. Res. Commun. 231:397-396(1997).
 CC -1- FUNCTION: MAY PLAY A ROLE IN NEGATIVE REGULATION OF APOPTOTIC
 CASCADES DURING OOGENESIS AND EMBRYOGENESIS, AND IN
 DEVELOPMENTALLY MATURED TISSUES SUCH AS BRAIN TISSUE.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
 CC -1- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN OVARIES AND TO A LESSER
 EXTENT IN TESTIS, INTESTINE, KIDNEY, BRAIN, EYE AND LUNG. WEAK
 EXPRESSION IN LIVER AND HEART. PRESENT IN TRACE AMOUNTS IN
 SKELETAL MUSCLE.
 CC -1- DEVELOPMENTAL STAGE: ABUNDANT IN EARLY STAGES OF OOGENESIS. THE
 EXPRESSION IS RAPIDLY REDUCED BETWEEN MEIOTIC MATURATION AND
 FERTILIZATION STAGES.
 CC -1- SIMILARITY: BELONGS TO THE PRESENTILIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D84427; BAA19570.1; -
 CC DR MEROPS: A22.001; -
 CC DR InterPro: IPR001108; -
 CC DR Pfam: PF01080; Presentilin; 1.
 CC DR PRINTS: PRO1072; PRESENTILIN.

KW Transmembrane; Glycoprotein.
 FT TRANSMEM 48 68 POTENTIAL.
 FT TRANSMEM 99 119 POTENTIAL.
 FT TRANSMEM 127 147 POTENTIAL.
 FT TRANSMEM 161 181 POTENTIAL.
 FT TRANSMEM 184 204 POTENTIAL.
 FT TRANSMEM 210 230 POTENTIAL.
 FT TRANSMEM 247 267 POTENTIAL.
 FT TRANSMEM 374 394 POTENTIAL.
 FT TRANSMEM 399 419 POTENTIAL.
 FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 371 371 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 433 AA: 48301 MW: 71CCE3F6BB9C0AF CRC64:

Query Match Best Local Similarity 92.5%; Score 37; DB 1; Length 433;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 LVNMAEGD 8
 |||||
 Db 261 LVNMAEGD 268

RESULT 7
 ID GYRA_MYCSM STANDARD: PRT: 842 AA.
 AC P48354;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE DNA GYRASE SUBUNIT A (EC 5.99.1.3).
 GN GYRA.
 OS Mycobacterium smegmatis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1772;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 700084 / MC(2)155;
 RX MEDLINE=97032832; PubMed=8878580;
 RA Revel-Viravau V., Truong Q.C., Moreau N., Jarlier V., Sougakoff W.;
 RT "Sequence analysis, purification, and study of inhibition by 4-
 quinolones of the DNA gyrase from Mycobacterium smegmatis.";
 RL Anticancerob. Agents Chemother. 40:2054-2061(1996).
 CC -1- FUNCTION: DNA GYRASE NEGATIVELY SUPERCOILS CLOSED CIRCULAR DOUBLE-
 STRANDED DNA IN AN ATP-DEPENDENT MANNER AND ALSO CATALYZES THE
 INTERCONVERSION OF OTHER TOPOLOGICAL ISOMERS OF DOUBLE-STRANDED
 DNA RINGS, INCLUDING CATENANES AND KNOTTED RINGS.
 CC -1- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
 OF DOUBLE-STRANDED DNA.
 CC -1- SUBUNIT: MADE UP OF TWO CHAINS. THE A CHAIN IS RESPONSIBLE FOR DNA
 BREAKAGE AND REJOINING; THE B CHAIN CATALYZES ATP HYDROLYSIS. THE
 ENZYME FORMS AN A2B2 TETRAMER.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X94224; CAA63918.1; -
 CC DR InterPro: IPR002205; -
 CC DR Pfam: PF00521; DNA_topoisomIV; 1.
 CC DR Topoisomerase; Isomerase; ATP-binding; Antibiotic resistance.
 CC ACT_SITE 130 130 DNA_CLEAVAGE (BY SIMILARITY).
 SQ SEQUENCE 842 AA: 93183 MW: 50A5DBE80ED45DD CRC64;

Query Match Best Local Similarity 85.0%; Score 34; DB 1; Length 842;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVNMAEGD 8
1:1:1111
DB 804 LVNMAEGD 811

RESULT 8
GSPD_VIBCH STANDARD: PRT: 674 AA.
ID GSPD_VIBCH
AC P45779;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE GENERAL SECRETION PATHWAY PROTEIN D PRECURSOR (CHOLERA TOXIN SECRETION
DE PROTEIN ERSD)
DE EPSP OR VC2733.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-EL TOR TRH7000;
RA Overbye L.J.;
RT "Organization of the general secretion pathway genes in Vibrio
RT cholerae."
RL Thesis (1994), Michigan State University / East Lansing, U.S.A.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gali S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Drygoi I., Sellers P.,
RA McDonald S.L., Uitterback T., Fleischmann R.D., Nieman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
CC -1- FUNCTION: REQUIRED FOR SECRETION OF CHOLERA TOXIN THROUGH THE
CC -1- OUTER MEMBRANE.
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE EXED/OUTD/PULD/XPSD FAMILY.
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT (POSSIBLY NATURAL) IN POSITION 55.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L33796; AAA58785.1;
DR EMBL: AE004338; -, NOT_ANNOTATED_CDS.
DR TIGR: VC2733; -;
DR InterPro: IPR00016; -;
DR InterPro: IPR001775; -;
DR Pfam: PF00263; Bac.GSPProteins; 1.
DR PRINTS: PR00811; BACTERIALGSPD.
DR PROSITE: PS00875; T2SP_D; 1.
KM Transport; Outer membrane; Signal.
FT SIGNAL 1 24
FT CHAIN 25 674
FT CONFLICT 89 89
FT CONFLICT 144 144
FT CONFLICT 144 144
SQ SEQUENCE 674 AA; 73469 MW; 3D77B891A59E6223 CRC64;

Query Match 80.0%; Score 32; DB 1; Length 674;

Best Local Similarity 75.0%; Pred. No. 20;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LVNMAEGD 8
1:1:1111
DB 350 LVNMAEGD 357

RESULT 9
ROH3_RAT STANDARD: PRT: 317 AA.
ID ROH3_RAT
AC P55006;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE RETINOL DEHYDROGENASE TYPE III (EC 1.1.1.105) (RODH III).
DE Rattus norvegicus (Rat).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96194805; PubMed=8647450;
RA Chai X., Zhai Y., Napoli J.L.;
RT "Cloning of a rat cDNA encoding retinol dehydrogenase isozyme type
RT III."
RL Gene 169:219-222(1996).
CC -1- FUNCTION: ACTS ON RETINOL BOUND ON CELLULAR RETINOL-BINDING
CC -1- PROTEIN (CRBP).
CC -1- CATALYTIC ACTIVITY: RETINOL + NAD(+) = RETINAL + NADH.
CC -1- PATHWAY: CATALYZES THE PRIMARY AND RATE-LIMITING STEP IN RETINOIC
CC ACID SYNTHESIS.
CC -1- SUBCELLULAR LOCATION: MICROSOMAL.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SCR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U33501; AAB07995.1; -;
DR HSP: P14061; 1PDV.
DR InterPro: IPR002198; -;
DR Pfam: PF00106; adh_short; 1.
DR PRINTS: PR00080; SDRFAMILY.
DR PROSITE: PS00061; ADH_SHORT; 1.
KW Oxidoreductase; NADP; NAD; Microsome.
KW NP_BIND 33
FT NP_BIND 176 176
FT ACT_SITE 176 176
SQ SEQUENCE 317 AA; 35736 MW; 8193672B8576897A CRC64;

Query Match 77.5%; Score 31; DB 1; Length 317;
Best Local Similarity 75.0%; Pred. No. 15;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 LVNMAEGD 8
1:1:1111
DB 253 LVNMAEGD 260

RESULT 10
SAHL_THEMA STANDARD: PRT: 404 AA.
ID SAHL_THEMA
AC O51933;
DT 15-DEC-1998 (Rel. 37, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ADENOSYLHOMOCYSTEINASE (EC 3.3.1.1) (S-ADENOSYL-L-HOMOCYSTEINE


```

DE HYDROLASE) (ADHOCYASE).
GN AHCY OR TM0172.
OS Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=98101466; PubMed=9440516;
RA de la Tour C.B., Portemer C., Kalloum H., Duguet M.;
RT *Reverse gyrase from the hyperthermophilic bacterium Thermotoga
RL maritima: properties and gene structure."
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT *Evidence for lateral gene transfer between Archaea and Bacteria from
RL genome sequence of Thermotoga maritima."
CC -1- CATALYTIC ACTIVITY: S-ADENOSYL-L-HOMOCYSTEINE + H(2)O =
CC -1- ADENOSINE + L-HOMOCYSTEINE.
CC -1- COFACTOR: NAD (BY SIMILARITY).
CC -1- PATHWAY: ACTIVATED METHYL CYCLE.
CC -1- SIMILARITY: BELONGS TO THE ADENOSYLHOMOCYSTEINASE FAMILY.
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: AF013268; AAC01562.1; -
CC DR EMBL: AE001702; AAD35265.1; -
CC DR TIGR: TM0172; -
CC DR InterPro: IPR000043; -
CC DR Pfam: PF00670; AdhCysase_1;
CC DR PROSITE: PS00738; ADHOCYASE_1; 1.
CC DR PROSITE: PS00739; ADHOCYASE_2; 1.
CC KM Hydrolase; NAD: One-carbon metabolism.
CC FT NP_BIND 197 228 NAD (POTENTIAL).
CC FT CONFLICT 44 44 A -> R (IN REF. 1).
CC FT CONFLICT 51 51 I -> V (IN REF. 1).
CC SO SEQUENCE 404 AA; 44849 MW; D20DB3BE02826148 CRC64;

Query Match 77.5%; Score 31; DB 1; Length 404;
Best Local Similarity 75.0%; Pred. NO. 19;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93209371; PubMed=8458428;
RA Ito M., Yasui A., Komamine A.;
RT "Precise mapping and molecular characterization of the MPT1 gene
RT involved in import of a fusion protein into mitochondria in
RL Saccharomyces cerevisiae."
RL PDBS Lett. 320:125-129(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Bowman S., Churcher C., Barrell B.G., Rajandream M.A.;
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN MITOCHONDRION PROTEIN IMPORT. BUT IS
CC NOT AN ESSENTIAL COMPONENT IN NORMAL GROWTH CONDITIONS.
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: S57517; AAB26005.1; -
CC DR EMBL: Z38114; CA86225.1; -
CC DR PIR: S32405; S32405.
CC DR PIR: S48336; S48336.
CC DR SGD: S00004527; MPT1.
CC KM Protein transport; Mitochondrion.
CC SO SEQUENCE 392 AA; 44996 MW; 45A775F8C1DA3E29 CRC64;

Query Match 75.0%; Score 30; DB 1; Length 392;
Best Local Similarity 75.0%; Pred. NO. 31;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

OY 1 LVNMAEGD 8
DB 327 LVNMAEGD 334

RESULT 11
MFTL YEAST STANDARD; PRT; 392 AA.
AC P33441;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE MITOCHONDRIAL FUSION TARGET PROTEIN.
GN MFT1 OR YML062C.
OS Saccharomyces cerevisiae (Baker's yeast).

```

```

OY 1 LVNMAEGD 8
DB 251 LVNMAEGD 258

RESULT 12
ARGJ_BACSU STANDARD; PRT; 406 AA.
AC P36843;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ARGININE BIOSYNTHESIS BIFUNCTIONAL PROTEIN ARGJ (INCLUDES: GLUTAMATE
DE N-ACETYLTRANSFERASE (EC 2.3.1.35) (ORNITHINE ACETYLTRANSFERASE)
DE (ORNITHINE TRANSACETYLASE) (OATASE); AMINO-ACID ACETYLTRANSFERASE
DE (EC 2.3.1.1) (N-ACETYLGLOUTAMATE SYNTHASE) (AAS)).
GN ARGJ.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=94297722; PubMed=8025667;
RA O'Reilly M., Devine K.M.;
RT "Sequence and analysis of the citrulline biosynthetic operon argC-F
RT from Bacillus subtilis."
RL Microbiology 140:1023-1025(1994).
RN [2]
RP SEQUENCE OF 1-80 FROM N.A.
RC STRAIN=168;
RX MEDLINE=98015415; PubMed=9353931;

```

RA Medina N., Vannier F., Roche B., Autret S., Levine A., Seror S.J.;
 RT "Sequencing of regions downstream of adda (98 degrees) and c1cg (289
 RL Microbiology 143:3305-3308(1997)."
 CC -1- CATALYTIC ACTIVITY: N2-ACETYL-L-ORNITHINE + L-GLUTAMATE =
 CC L-ORNITHINE + N-ACETYL-L-GLUTAMATE.
 CC -1- CATALYTIC ACTIVITY: ACETYL-COA + L-GLUTAMATE = COA + N-ACETYL-
 CC L-GLUTAMATE.
 CC -1- PATHWAY: FIRST AND FIFTH STEPS IN ARGININE BIOSYNTHESIS.
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOSOLASMIC (PROBABLE).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; 226919; CAAB1544.1; -
 CC DR EMBL; 279580; CAB01843.1; -
 CC DR EMBL; Y09476; CAAY0639.1; -
 CC DR EMBL; 299109; CAB12961.1; -
 CC DR EMBL; 299110; CAB12977.1; -
 CC DR PIR; S38429; S38429.
 CC DR Subtilist; BG10192; argJ.
 CC DR InterPro; IPR002813; -
 CC DR Pfam; PF01960; ArgJ; 1.
 CC DR Arginine biosynthesis; Transferase; Acyltransferase;
 CC KW Multifunctional enzyme.
 CC SO SEQUENCE 406 AA; 43364 MW; D49C9192055D0571 CRC64;

Query Match 75.0%; Score 30; DB 1; Length 406;
 Best Local Similarity 62.5%; Pred. No. 32;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 LVNMGD 8
 DB 376 VIKMGD 383

RESULT 13
 GR75-CRIGR STANDARD; PRT; 679 AA.
 ID GR75-CRIGR
 AC 035501;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE MITOCHONDRIAL STRESS-70 PROTEIN PRECURSOR (75 KDA GLUCOSE REGULATED
 DE PROTEIN) (GRP 75).
 GN HSPA9.
 OS Crictetus griseus (Chinese hamster).
 OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Crictetus
 OC NCBI_Taxid=10029;
 OX RN
 RP SEQUENCE FROM N.A.
 RA MEDLINE-97405843; PubMed-9260887;
 RA Singh B., Solys B.J., Wu Z.C., Patel H.V., Freeman K.B., Gupta R.S.;
 RT "Cloning and some novel characteristics of mitochondrial Hsp70 from
 RT Chinese hamster cells".
 RL Exp. Cell Res. 234:205-216(1997).
 CC -1- FUNCTION: IMPLICATED IN THE CONTROL OF CELL PROLIFERATION AND
 CC CELLULAR AGING. MAY ALSO ACT AS A CHAPERONE (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL.
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; U92313; AAB62091.1; -
 CC DR HSSP; P04475; IDKZ.
 CC DR InterPro; IPR001023; -
 CC DR Pfam; PF00012; HSP70_1.
 CC DR PRINTS; PRO0301; HEATSHOCK70.
 CC DR PROSITE; PS00297; HSP70_1; 1.
 CC DR PROSITE; PS00329; HSP70_2; 1.
 CC DR PROSITE; PS01036; HSP70_3; 1.
 CC KW ATP-binding; Mitochondrion; Transit peptide.
 CC FT TRANSIT 1 46 MITOCHONDRIAL (BY SIMILARITY).
 CC FT CHAIN 47 679 MITOCHONDRIAL STRESS-70 PROTEIN.
 CC SO SEQUENCE 679 AA; 73730 MW; CCB62241B1ECAA70 CRC64;

Query Match 75.0%; Score 30; DB 1; Length 679;
 Best Local Similarity 100.0%; Pred. No. 56;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 VNMAEG 7
 DB 582 VNMAEG 587

RESULT 14
 GR75-HUMAN STANDARD; PRT; 679 AA.
 ID GR75-HUMAN
 AC P38646; P30036; P31932;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE MITOCHONDRIAL STRESS-70 PROTEIN PRECURSOR (75 KDA GLUCOSE REGULATED
 DE PROTEIN) (GRP 75) (PEPTIDE-BINDING PROTEIN 74) (PBP74) (MORTALIN)
 DE (MOT).
 GN HSPA9B OR HSPA9 OR GRP75.
 OS Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_Taxid=9606;
 OX RN
 RP SEQUENCE FROM N.A.
 RA MEDLINE-95130547; PubMed-7829505;
 RA Bhattacharya T., Karnezis A.N., Murphy S.P., Hoang T., Freeman B.C.,
 RA Phillips B., Morimoto R.I.;
 RT "Cloning and subcellular localization of human mitochondrial hsp70.";
 RL J. Biol. Chem. 270:1705-1710(1995).
 RN [1]
 RP SEQUENCE FROM N.A.
 RA TISSUE-B-cell;
 RC MEDLINE-93268309; PubMed-7684501;
 RA Domancico S.Z., Denagel D.C., Dahlseid J.N., Green J.M.,
 RA Pierce S.K.;
 RT "Cloning of the gene encoding peptide-binding protein 74 shows that
 RT it is a new member of the heat shock protein 70 family.";
 RL Mol. Cell. Biol. 13:3598-3610(1993).
 RN [3]
 RP SEQUENCE OF 47-56.
 RC TISSUE-Liver;
 RX MEDLINE-93162045; PubMed-1286669;
 RA Hochstrasser D.F., Frutiger S., Paquet N., Bairoch A., Ravier F.,
 RA Pasquali C., Sanchez J.-C., Tissot J.-D., Bjellqvist B., Vargis R.,
 RA Appel R.D., Hughes G.J.;
 RT "Human liver protein map: a reference database established by
 RT microsequencing and gel comparison.";
 RL Electrophoresis 13:992-1001(1992).
 RN [4]
 RP REVISIONS.
 RC TISSUE-Liver;
 RX MEDLINE-94147969; PubMed-8313870;

RA Hughes G.J., Frutiger S., Paquet N., Pasquall C., Sanchez J.-C.,
 RA Tissot J.-D., Bairoch A., Appel R.D., Hochstrasser D.F.;
 RT "Human liver protein map: update 1993."; <http://www.isb-sib.ch/announce/>
 RL Electrophoresis 14:1216-1222(1993).
 RN [5]
 RP SEQUENCE OF 47-68.
 RC TISSUE=Colon carcinoma;
 RX MEDLINE=97295306; PubMed=9150948;
 RA JI H., Reid G.E., Moritz R.L., Eddes J.S., Burgess A.W., Simpson R.J.;
 RT "A two-dimensional gel database of human colon carcinoma proteins."; <http://www.isb-sib.ch/announce/>
 RL Electrophoresis 18:605-613(1997).
 CC -1- FUNCTION: IMPLICATED IN THE CONTROL OF CELL PROLIFERATION AND
 CC CELLULAR AGING. MAY ALSO ACT AS A CHAPERONE.
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L15189; AAA67526.1; -;
 DR EMBL: L11066; -; NOT_ANNOTATED_CDS.
 DR PIR: B48127; B48127.
 DR HSSP: P04475; IDXX.
 DR SWISS-2DPAGE: P38646; HUMAN.
 DR HSC-2DPAGE: P38646; HUMAN.
 DR MIM: 600546; -;
 DR InterPro: IPR001023; -;
 DR Pfam: PF00012; HSP70; 1.
 DR PRINTS: PR00301; HEATSHOCK70.
 DR PROSITE: PS00297; HSP70.1; 1.
 DR PROSITE: PS00329; HSP70.2; 1.
 DR PROSITE: PS01036; HSP70_3; 1.
 DR ATP-binding; Mitochondrion; Transl. peptide.
 KW ATP-binding; Mitochondrion; Transl. peptide.
 FT TRANSIT 1 46 MITOCHONDRION.
 FT CHAIN 47 679 MITOCHONDRIAL STRESS-70 PROTEIN.
 FT CONFLICT 540 540 R -> G (IN REF. 2).
 FT SEQUENCE 679 AA: 73779 MW: 908D9B606D57653 CRC64;
 SQ
 Query Match 75.0%; Score 30; DB 1; Length 679;
 Best Local Similarity 100.0%; Pred. No. 56;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 VNMAEG 7
 DB 582 VNMAEG 587
 RESULT 15
 GR5_MOUSE STANDARD: PRT; 679 AA.
 ID GR5_MOUSE
 AC P38647;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE MITOCHONDRIAL STRESS-70 PROTEIN PRECURSOR (75 KDA GLUCOSE REGULATED
 DE PROTEIN) (GRP 75) (PEPTIDE-BINDING PROTEIN 74) (P66 MOT)
 DE (MORTALIN)
 GN HSPA9A OR HSPA9 OR GRP75 OR HSP74.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CD-1-ICR; TISSUE=Embryonic fibroblast;
 RX MEDLINE=93203261; PubMed=8454632;
 RA Madhwa R., Kaul S.C., Ikawa Y., Sugimoto Y.;

RT "Identification of a novel member of mouse hsp70 family. Its
 RT association with cellular mortal phenotype."; <http://www.isb-sib.ch/announce/>
 RL J. Biol. Chem. 268:6615-6621(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CD-1-ICR; TISSUE=Embryonic fibroblast;
 RX MEDLINE=94042962; PubMed=7693662;
 RA Madhwa R., Kaul S.C., Sugimoto Y., Mitsui Y.;
 RT "Induction of cellular senescence by transfection of cytosolic
 RT mortalin cDNA in NIH 3T3 cells."; <http://www.isb-sib.ch/announce/>
 RL J. Biol. Chem. 268:22239-22242(1993).
 RN [3]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=B-cell;
 RX MEDLINE=93268309; PubMed=7684501;
 RA Domonic S.Z., Denagel D.C., Dahlseid J.N., Green J.M., Pierce S.K.;
 RT "Cloning of the gene encoding peptide-binding protein 74 shows that
 RT it is a new member of the heat shock protein 70 family."; <http://www.isb-sib.ch/announce/>
 RL Mol. Cell. Biol. 13:3598-3610(1993).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BALB/C; TISSUE=Liver;
 RX MEDLINE=94085585; PubMed=8262211;
 RA Michikawa Y., Baba T., Arai Y., Sakakura T., Kusakabe M.;
 RT "Structure and organization of the gene encoding a mouse
 RT mitochondrial stress-70 protein."; <http://www.isb-sib.ch/announce/>
 RL FEBS Lett. 336:27-33(1993).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN-C3H/HEJ; TISSUE=Kidney;
 RX MEDLINE=94029998; PubMed=7692847;
 RA Michikawa Y., Baba T., Arai Y., Sakakura T., Tanaka M., Kusakabe M.;
 RT "Antigenic protein specific for C3H strain mouse is a mitochondrial
 RT stress-70 protein."; <http://www.isb-sib.ch/announce/>
 RL Biochem. Biophys. Res. Commun. 196:223-232(1993).
 RN [6]
 RP SEQUENCE OF 47-70.
 RC TISSUE=fibroblast;
 RX MEDLINE=95009907; PubMed=7523108;
 RA Merrick B.A., Patterson R.M., Wichter L.D., He C., Selkirk J.K.;
 RT "Separation and sequencing of familial and novel murine proteins
 RT using preparative two-dimensional gel electrophoresis."; <http://www.isb-sib.ch/announce/>
 RL Electrophoresis 15:735-743(1994).
 RN [7]
 RP SUBCELLULAR LOCATION.
 RC MEDLINE=95170122; PubMed=7865888;
 RA Dahlseid J.N., Lill R., Green J.M., Xu X., Olu Y., Pierce S.K.;
 RT "PBP74, a new member of the mammalian 70-kda heat shock protein
 RT family, is a mitochondrial protein."; <http://www.isb-sib.ch/announce/>
 RL Mol. Biol. Cell 5:1265-1275(1994).
 CC -1- FUNCTION: IMPLICATED IN THE CONTROL OF CELL PROLIFERATION AND
 CC CELLULAR AGING. MAY ALSO ACT AS A CHAPERONE.
 CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL.
 CC -1- TISSUE SPECIFICITY: FOUND IN ALL THE CELL TYPES EXAMINED.
 CC -1- INDUCTION: NOT INDUCED BY HEAT SHOCK, INSTEAD PROTEIN LEVEL IS
 CC DECREASED.
 CC -1- POLYMORPHISM: TWO FORMS OF THE PROTEIN HAVE BEEN FOUND, MOT-1,
 CC FOUND IN MORTAL CELLS AND MOT-2, FOUND IN IMMORTAL CELLS.
 CC THE SEQUENCE OF MOT-1 IS SHOWN HERE.
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D11089; BAA01862.1; -;
 DR EMBL: I06886; -; NOT_ANNOTATED_CDS.
 DR EMBL: D17666; BAA04548.1; -;
 DR EMBL: D17662; BAA04548.1; JOINED.

DR EMBL; D17655; BAA04548.1; JOINED.
DR EMBL; D17657; BAA04548.1; JOINED.
DR EMBL; D17658; BAA04548.1; JOINED.
DR EMBL; D17659; BAA04548.1; JOINED.
DR EMBL; D17660; BAA04548.1; JOINED.
DR EMBL; D17663; BAA04548.1; JOINED.
DR EMBL; D17664; BAA04548.1; JOINED.
DR EMBL; D17665; BAA04548.1; JOINED.
DR EMBL; D17556; BAA04493.1; -
DR PIR; A45497; A45497.
DR PIR; A48698; A48698.
DR PIR; A48698; A48698.
DR HSSP; P04475; 1DKX.
DR SWISS-2DPAGE; P38647; MOUSE.
DR MGD; MGI:96245; Hspa9a.
DR InterPro; IPR001023; -
DR Pfam; PF00012; HSP70; 1.
DR PRINTS; PR00301; HEATSHOCK70.
DR PROSITE; PS00297; HSP70_1; 1.
DR PROSITE; PS00329; HSP70_2; 1.
DR PROSITE; PS01036; HSP70_3; 1.
KW ATP-binding; Mitochondrion; Transit peptide.
KW TRANSIT 1
FT CHAIN 1 46 MITOCHONDRION.
FT VARIANT 618 618 V -> M (IN MOT-2, PERINUCLEAR FORM).
FT VARIANT 624 624 R -> G (IN MOT-2, PERINUCLEAR FORM).
FT CONFLICT 5 5 S -> T (IN REF. 3).
FT CONFLICT 106 106 K -> R (IN REF. 3).
FT CONFLICT 123 123 I -> L (IN REF. 3, 4 AND 5).
FT CONFLICT 203 203 Q -> D (IN REF. 2).
FT CONFLICT 522 522 F -> S (IN REF. 3).
SQ SEQUENCE 679 AA: 73528 MW: FE0F6C627492A4AD CRC64;

Query Match 75.0%; Score 30; DB 1; Length 679;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VNMAEG 7
|||||
Db 582 VNMAEG 587

-- Search completed: September 6, 2001, 16:51:08
Job time: 812 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:45:50 ; Search time 134.15 seconds
(without alignments)
4.543 Million cell updates/sec

Title: US-09-603-713-9

Perfect score: 40

Sequence: 1 LVNMAECD 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR6:
2: PIR2:
3: PIR3:
4: PIR4:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	100.0	374	2 S63684	presenilin 1, sp1
2	40	100.0	463	2 S63683	presenilin I-463 -
3	40	100.0	463	2 JCS081	presenilin 1 prote
4	40	100.0	467	2 S58396	presenilin 1, sp1
5	40	100.0	467	2 JCS080	presenilin 1 prote
6	40	100.0	467	2 I78388	S182 protein - mou
7	37	92.5	433	2 JCS390	presenilin-alpha -
8	33	82.5	562	2 H70688	hypothetical prote
9	32	80.0	833	2 H81700	DNA gyrase, chain
10	32	80.0	836	2 F71546	probable DNA gyras
11	32	80.0	3016	2 S77300	hypothetical prote
12	31	77.5	184	2 H84780	probable AP2 domai
13	31	77.5	334	2 T51245	COP1-interacting p
14	31	77.5	354	2 T35599	conserved hypotet
15	31	77.5	404	2 B72409	adenosylhomocystei
16	31	77.5	431	2 A72277	amidophosphoribosy
17	31	77.5	436	2 E81103	conserved hypotet
18	31	77.5	472	2 T30101	hypothetical prote
19	31	77.5	662	2 T01857	hypothetical prote
20	31	77.5	765	2 G64502	hypothetical prote
21	31	77.5	1420	2 T02644	ABC-type transport
22	30	75.0	188	2 A86260	protein T12C24.16
23	30	75.0	216	2 JEO297	DRE/CRT-binding pr
24	30	75.0	216	2 T51830	transcription fact
25	30	75.0	225	2 T03125	hypothetical prote
26	30	75.0	314	2 D85294	transcription acti
27	30	75.0	314	2 T05799	transcription acti
28	30	75.0	335	2 G81280	holliday junction
29	30	75.0	392	2 S32405	MFT1 protein - yea

30	30	75.0	406	2 I40373	ornithine acetyltr
31	30	75.0	524	2 E75574	aldehyde dehydrog
32	30	75.0	601	2 T45664	hypothetical prote
33	30	75.0	679	1 B48127	dnak-type molecule
34	30	75.0	679	1 I55581	dnak-type molecule
35	30	75.0	679	2 A48127	dnak-type molecule
36	30	75.0	723	1 JN0531	p-aminobenzoic aci
37	30	75.0	838	2 D70698	probable gyra prot
38	30	75.0	859	2 S49046	luxO protein - Vib
39	30	75.0	1273	2 T10006	DNA topoisomerase
40	30	75.0	1598	2 S68967	Tyb protein - yea
41	30	75.0	1755	2 S68949	Tyb protein - yea
42	30	75.0	1770	2 S68953	Tyb protein - yea
43	30	75.0	1770	2 S68948	Tyb protein - yea
44	30	75.0	1770	2 S70233	Tyb protein - yea
45	30	75.0	1770	2 S70230	Tyb protein - yea

ALIGNMENTS

RESULT 1
S63684
presenilin 1, splice form 374 - human
N:Alternate names: Alzheimer's disease protein 3
C:Species: Homo sapiens (man)
C>Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 29-Sep-1999
C:Accession: S63684
R:Sahara, N.; Yahagi, Y.; Takagi, H.; Kondo, T.; Okochi, M.; Usami, M.; Shirasawa, T.
FEBS Lett. 381, 7-11, 1996
A:Title: Identification and characterization of presenilin I-467, I-463 and I-374.
A:Reference number: S63683; MUID:96193901
A:Accession: S63684
A>Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-374 <SAH>
A:Cross-references: EMBL:U40380; NID:g1244639; PIDN:AAB05895.1; PID:g1244640
C:Genetics:
A:Gene: GDB:PSEN1; AD3; FAD; S182; PS1
A:Cross-references: GDB:135682; OMIM:104311
A:Map position: 14q24.3-14q24.3
C:Superfamily: presenilin

Query Match 100.0%; Score 40; DB 2; Length 374;
Best Local Similarity 100.0%; Pred. No. 0.52;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVNMAECD 8
Db 291 LVNMAECD 298

RESULT 2
S63683
presenilin I-463 - human
C:Species: Homo sapiens (man)
C>Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 29-Sep-1999
C:Accession: S63683
R:Sahara, N.; Yahagi, Y.; Takagi, H.; Kondo, T.; Okochi, M.; Usami, M.; Shirasawa, T.
FEBS Lett. 381, 7-11, 1996
A:Title: Identification and characterization of presenilin I-467, I-463 and I-374.
A:Reference number: S63683; MUID:96193901
A:Accession: S63683
A>Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-463 <SAH>
A:Cross-references: EMBL:U40379; NID:g1244637; PIDN:AAB05894.1; PID:g1244638
C:Superfamily: presenilin

Query Match 100.0%; Score 40; DB 2; Length 463;
Best Local Similarity 100.0%; Pred. No. 0.65;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVNMAEGD 8
|||||||

Db 291 LVNMAEGD 298

RESULT 3

JC5081
presenilin 1 protein isoform 463 - lesser mouse lemur
C:Species: Microcebus murinus (lesser mouse lemur)
C:Date: 31-Jan-1997 #sequence_revision 31-Jan-1997 #text_change 13-Sep-1998
C:Accession: JC5081
R:Calenda, A.; Mestre-Frances, N.; Czech, C.; Pradier, L.; Petter, A.; Bons, N.; Bellis, Biochem. Biophys. Res. Commun. 228, 430-439, 1996
A:Title: Molecular cloning, sequencing, and brain expression of the presenilin 1 gene in A:Reference number: JC5080; MUID:97079199
A:Contents: brain
A:Accession: JC5081
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-463 <CAL>
A:Cross-references: EMBL:271333
C:Comment: This protein is an intermembrane protein with seven transmembrane domains. It C:Genetics:
A:Gene: psi
A:Map position: 14
C:Superfamily: presenilin
C:Keywords: transmembrane protein
F:78-96/Domain: transmembrane #status predicted <TM1>
F:129-150/Domain: transmembrane #status predicted <TM2>
F:160-181/Domain: transmembrane #status predicted <TM3>
F:191-209/Domain: transmembrane #status predicted <TM4>
F:217-234/Domain: transmembrane #status predicted <TM5>
F:240-257/Domain: transmembrane #status predicted <TM6>
F:404-424/Domain: transmembrane #status predicted <TM7>

Query Match 100.0%; Score 40; DB 2; Length 463;
Best Local Similarity 100.0%; Pred. No. 0.65;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVNMAEGD 8
|||||||

Db 291 LVNMAEGD 298

RESULT 4

S58396
presenilin 1, splice form 467 - human
N:Alternate names: Alzheimer's disease protein 3; protein S182
C:Species: Homo sapiens (man)
C:Date: 29-Jan-1998 #sequence_revision 13-Feb-1998 #text_change 29-Sep-1999
C:Accession: S58396; S71401; S71402
R:Sherrington, R.; Rogae, E.; Liang, Y.; Rogae, E.A.; Levesque, G.; Ikeda, M.; Chi, E.; Romenes, J.M.; St George-Hyslop, P.H.
A:Title: Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's A:Reference number: I58095; MUID:95319502
A:Accession: S58396
A:Molecule type: mRNA
A:Residues: 1-467 <SHE>
A:Cross-references: EMBL:LA2110; NID:9904118; PIDN:AAB46416.1; PID:9904119
R:Vidal, R.; Ghiso, J.; Wisniewski, T.; Frangione, B.
FBS Lett. 393, 19-23, 1996
A:Title: Alzheimer's presenilin 1 gene expression in platelets and megakaryocytes. Ident A:Reference number: S71401; MUID:96397521
A:Accession: S71401
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 24-32;234-256,290-292;316-317,376-379 <VIN>
A:Experimental source: Dami megakaryotic cell line (ATCC CRL-9792) and platelets

C:Genetics:
A:Gene: GDB:PSEN1; AD3; FAD; S182; PS1
A:Cross-references: GDB:135682; OMIM:104311
A:Map position: 14q24.3-14q24.3
C:Superfamily: presenilin
C:Keywords: alternative splicing; Alzheimer's disease; glycoprotein; transmembrane pr

F:82-100/Domain: transmembrane #status predicted <TM1>
F:133-154/Domain: transmembrane #status predicted <TM2>
F:164-185/Domain: transmembrane #status predicted <TM3>
F:195-213/Domain: transmembrane #status predicted <TM4>
F:221-238/Domain: transmembrane #status predicted <TM5>
F:241-264/Domain: transmembrane #status predicted <TM6>
F:281-301/Domain: transmembrane #status predicted <TM7>
F:408-428/Domain: transmembrane #status predicted <TM8>
F:433-453/Domain: transmembrane #status predicted <TM9>
F:279,405/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 40; DB 2; Length 467;
Best Local Similarity 100.0%; Pred. No. 0.65;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVNMAEGD 8
|||||||

Db 295 LVNMAEGD 302

RESULT 5

JC5080
presenilin 1 protein isoform 467 - lesser mouse lemur
C:Species: Microcebus murinus (lesser mouse lemur)
C:Date: 31-Jan-1997 #sequence_revision 31-Jan-1997 #text_change 20-Jun-2000
C:Accession: JC5080
R:Calenda, A.; Mestre-Frances, N.; Czech, C.; Pradier, L.; Petter, A.; Bons, N.; Bell Biochem. Biophys. Res. Commun. 228, 430-439, 1996
A:Title: Molecular cloning, sequencing, and brain expression of the presenilin 1 gene A:Reference number: JC5080; MUID:97079199
A:Accession: JC5080
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-467 <CAL>
A:Cross-references: EMBL:271333; NID:91707591; PIDN:CA95930.1; PID:91707592
A:Experimental source: brain
C:Comment: This protein is an integral membrane protein with seven transmembrane dom C:Genetics:
A:Gene: psi
A:Map position: 14
C:Superfamily: presenilin
C:Keywords: transmembrane protein
F:82-100/Domain: transmembrane #status predicted <TM1>
F:133-154/Domain: transmembrane #status predicted <TM2>
F:164-185/Domain: transmembrane #status predicted <TM3>
F:195-213/Domain: transmembrane #status predicted <TM4>
F:221-238/Domain: transmembrane #status predicted <TM5>
F:244-261/Domain: transmembrane #status predicted <TM6>
F:408-428/Domain: transmembrane #status predicted <TM7>

Query Match 100.0%; Score 40; DB 2; Length 467;
Best Local Similarity 100.0%; Pred. No. 0.66;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVNMAEGD 8
|||||||

Db 295 LVNMAEGD 302

RESULT 6

I78388
S182 protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 27-Feb-1997 #sequence_revision 27-Feb-1997 #text_change 29-Sep-1999
C:Accession: I78388

R.Sherrington, R.; Rogaev, E.I.; Liang, Y.; Rogaeva, E.A.; Levesque, G.; Ikeda, M.; Chi,
ero, I.; Pinessi, L.; Nee, L.; Chumakov, I.; Pollen, D.; Brookes, A.; Sansseau, P.; Pollin
E.; Romanens, J.M.; St George-Hyslop, P.H.

Nature 375, 754-760, 1995
A>Title: Cloning of a gene bearing missense mutations in early-onset familial Alzheimer'
A:Reference number: 158095; PMID:95319502

A:Accession: 178388

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-467 <RES>

A:Cross-references: GB:142177; NID:9904129; PIDN:AAC42094.1; PID:9304130

C:Superfamily: presenilin

Query Match 100.0%; Score 40; DB 2; Length 467;

Best Local Similarity 100.0%; Pred. No. 0.66;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LVNMAEGD 8

DB 295 LVNMAEGD 302

RESULT 7

JC5390

presenilin-alpha - African clawed frog

C:Species: Xenopus laevis (African clawed frog)

C>Date: 04-Jun-1997 #sequence_revision 18-Jul-1997 #text_change 20-Jun-2000

C:Accession: JC5390

R:Tsujimura, A.; Yasojima, K.; Hashimoto-Gotoh, T.

Biochem. Biophys. Res. Commun. 231, 392-396, 1997

A>Title: Cloning of Xenopus presenilin-alpha and -beta cDNAs and their differential exp

A:Reference number: JC5390; PMID:97223465

A:Accession: JC5390

A:Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-433 <TSU>

A:Cross-references: DDBJ:D84427; NID:91944353; PIDN:BA19570.1; PID:91944354

A:Experimental source: brain

C:Comment: This protein plays a role in negative regulation of apoptotic cascades during

C:Superfamily: presenilin

F:48-66/Domain: transmembrane #status predicted <TM1>

F:99-119/Domain: transmembrane #status predicted <TM2>

F:130-149/Domain: transmembrane #status predicted <TM3>

F:161-178/Domain: transmembrane #status predicted <TM4>

F:187-203/Domain: transmembrane #status predicted <TM5>

F:210-227/Domain: transmembrane #status predicted <TM6>

F:374-394/Domain: transmembrane #status predicted <TM7>

Query Match 92.5%; Score 37; DB 2; Length 433;

Best Local Similarity 87.5%; Pred. No. 2.8;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 LVNMAEGD 8

DB 261 LVNMAEGD 268

RESULT 8

H70688

hypothetical protein RV2797C - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: H70688

R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Nature 393, 537-544, 1998

A:Authors: Seares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.

A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; PMID:98295987

A:Accession: H70688

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-562 <COL>

A:Cross-references: GB:Z81331; GB:AL123456; NID:93261650; PIDN:CAB03649.1; PID:e27681

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: RV2797C

Query Match 82.5%; Score 33; DB 2; Length 562;

Best Local Similarity 85.7%; Pred. No. 28;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 VNMAGD 8

DB 159 VNMAGD 165

RESULT 9

H81700

DNA gyrase, chain A TC0461 [imported] - Chlamydia muridarum (strain N199)

C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn

C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-May-2000

C:Accession: H81700

R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke

, C.; Dodson, R.; Gwin, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe

Nucleic Acids Res. 28, 1397-1406, 2000

A>Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39

A:Reference number: AB1500; PMID:20150255

A:Accession: H81700

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-833 <DET>

A:Cross-references: GB:AE002314; GB:AE002160; NID:97190495; PIDN:AAF39312.1; PID:9719

A:Experimental source: strain N199 (MoPn)

C:Genetics:

A:Gene: TC0461

C:Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain A; phage T4 DNA topoisomeras

Query Match 80.0%; Score 32; DB 2; Length 833;

Best Local Similarity 75.0%; Pred. No. 71;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 LVNMAEGD 8

DB 795 LVNMAEGD 802

RESULT 10

F71546

probable DNA gyrase chain A - Chlamydia trachomatis (serotype D, strain UM3/Cx)

C:Species: Chlamydia trachomatis

C>Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 03-Dec-1999

C:Accession: F71546

R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche

Science 282, 754-759, 1998

A>Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t

A:Reference number: A71570; PMID:99000809

A:Accession: F71546

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-836 <ARN>

A:Cross-references: GB:AE001292; GB:AE001273; NID:93328586; PIDN:AAC67781.1; PID:9332

A:Experimental source: serotype D, strain UM-3/Cx

C:Genetics:

A:Gene: gyrA_1

C:Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain A; phage T4 DNA topoisomeras

F:1-238/Domain: phage T4 DNA topoisomerase (ATP-hydrolyzing) medium chain homology <T

Query Match 80.0%; Score 32; DB 2; Length 836;

Best Local Similarity 75.0%; Pred. No. 72;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 LVNMAEGD 8
 |||: |||
 Db 795 LVNLRGCD 802

RESULT 11

S77300 hypothetical protein slr1403 - *Synechocystis* sp. (strain PCC 6803)

C:Species: *Synechocystis* sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999

C:Accession: S77300

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;

O, K.; Okumura, S.; Shimpo, S.; Takuch, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*

S.

A:Reference number: S74322; MUID:97061201

A:Accession: S77300

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-3016 <KAN>

A:Cross-references: EMBL:D90907; GB:AB001339; NID:g1652618; PIDN:BA017634.1; PID:d101836

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match

Best Local Similarity 80.0%; Score 32; DB 2; Length 3016;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LVNMAEGD 8
 |||: |||
 Db 1659 LVNLRGCD 1666

RESULT 12

H84780 probable AP2 domain transcription factor [Imported] - *Arabidopsis thaliana*

C:Species: *Arabidopsis thaliana* (mouse-ear cross)

C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C:Accession: H84780

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shee, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Selzberg, S.L.; Fraser, C.M.; Venter, J.

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.

A:Reference number: A84420; MUID:20083487

A:Accession: H84780

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-184 <STO>

A:Cross-references: GB:AE002093; NID:g4581145; PIDN:AAD24629.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g36450

A:Map position: 2

Query Match

Best Local Similarity 77.5%; Score 31; DB 2; Length 184;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVNMAEG 7
 |||: |||
 Db 147 LVNMAEG 153

RESULT 13

T51245 COP1-interacting protein CIP8 [Imported] - *Arabidopsis thaliana*

C:Species: *Arabidopsis thaliana* (mouse-ear cross)

C:Date: 28-Jul-2000 #sequence_revision 28-Jul-2000 #text_change 28-Jul-2000

C:Accession: T51245
 R:Torii, K.U.; Stoop-Myer, C.D.; Okamoto, H.; Coleman, J.E.; Matsui, M.; Deng, X.W.

J. Biol. Chem. 274, 27674-27681, 1999

A:Title: The RING finger motif of photomorphogenic repressor COP1 specifically interacts

A:Reference number: 225339; MUID:99419050

A:Accession: T51245

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-334 <TOR>

A:Cross-references: EMBL:AF162150; PIDN:AAD56636.1

C:Genetics:

A:Gene: CIP8

A:Map position: 5

Query Match

Best Local Similarity 77.5%; Score 31; DB 2; Length 334;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LVNMAEGD 8
 |||: |||
 Db 213 LVNLRGCD 220

RESULT 14

T39599 conserved hypothetical protein SPBC16G5.07c - fission yeast (*Schizosaccharomyces pombe*)

C:Species: *Schizosaccharomyces pombe*

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 18-Feb-2000

C:Accession: T39599

R:Lyne, M.; Wood, V.; Rajandream, M.A.; Barrell, B.G.; Beck, A.; Reinhardt, R.

submitted to the EMBL Data Library, May 1998

A:Reference number: 221866

A:Accession: T39599

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-354 <LYN>

A:Cross-references: EMBL:AL023554; PIDN:CA019027.1; GSPDB:GN00067; SPDB:SPBC16G5.07c

A:Experimental source: strain 972h-; cosmid cl65

C:Genetics:

A:Gene: SPDB:SPBC16G5.07c

A:Map position: 2

A:Antions: 72/2; 265/2

C:Superfamily: erythrocyte band 7 integral membrane protein

Query Match

Best Local Similarity 77.5%; Score 31; DB 2; Length 354;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 VNMAGECD 8
 |||: |||
 Db 230 INVAGECD 236

RESULT 15

B72409 adenosylhomocysteinease - *Thermotoga maritima* (strain MSB8)

C:Species: *Thermotoga maritima*

C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000

C:Accession: B72409

R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic

Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,

C.M.

Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome

A:Reference number: A72200; MUID:99287316

A:Accession: B72409

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-404 <ARN>

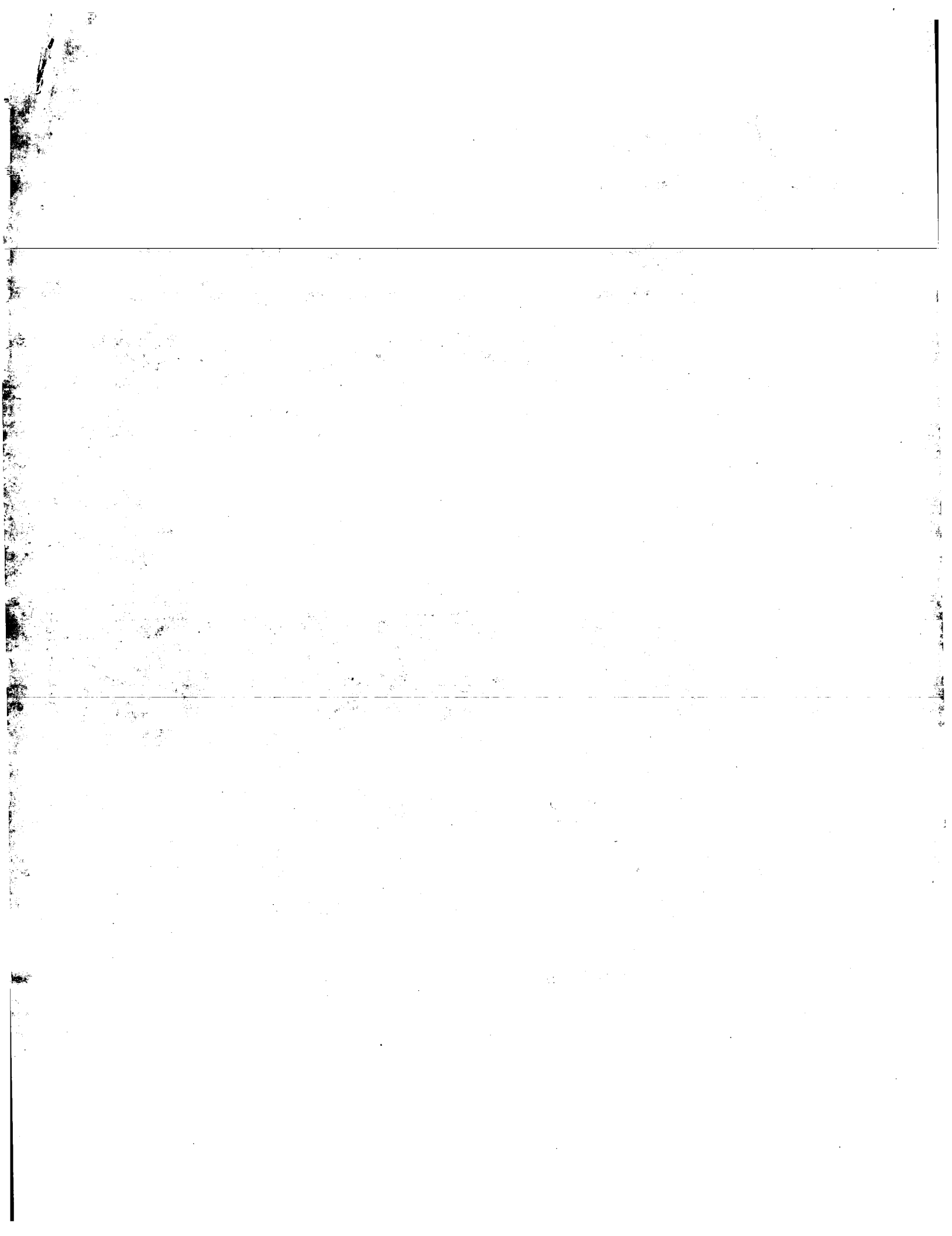
A:Cross-references: GB:AE001702; GB:AE000512; NID:g4980662; PIDN:AAD35265.1; PID:g498

A:Experimental source: strain MSB8

C:Genetics:
A:Gene: TM0172
C:Superfamily: adenosylhomocysteinase

Query Match 77.5%; Score 31; DB 2; Length 404;
Best Local Similarity 75.0%; Pred. No. 55;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 LVNMAEGD 8
|||:|
Db 327 LVNLAAGD 334

Search completed: September 6, 2001, 16:45:51
Job time: 495 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:39:36 ; Search time 113.12 Seconds
(without alignments)
1.456 Million cell updates/sec

Title: US-09-603-713-9

Perfect score: 40

Sequence: 1 LVNNAEGD 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCRTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfilest1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	100.0	467	2	US-08-875-972-4
2	40	100.0	462	3	US-08-788-231A-15
3	40	100.0	463	2	US-08-670-479-18
4	40	100.0	463	3	US-08-670-964-4
5	40	100.0	463	3	US-08-888-077A-4
6	40	100.0	465	3	US-08-788-231A-17
7	40	100.0	467	2	US-08-967-101-2
8	40	100.0	467	2	US-08-967-101-4
9	40	100.0	467	2	US-08-967-101-134
10	40	100.0	467	2	US-08-592-541-2
11	40	100.0	467	2	US-08-592-541-4
12	40	100.0	467	2	US-08-592-541-134
13	40	100.0	467	3	US-08-923-454A-10
14	40	100.0	467	3	US-08-670-964-2
15	40	100.0	467	3	US-08-888-077A-2
16	40	100.0	467	3	US-08-888-077A-17
17	40	100.0	467	3	US-08-124-698-2
18	40	100.0	467	3	US-09-124-698-4
19	40	100.0	467	3	US-09-124-698-134
20	40	100.0	467	4	US-09-127-480-2
21	40	100.0	467	4	US-09-127-480-4
22	40	100.0	467	4	US-09-127-480-134
23	40	100.0	467	4	US-08-496-841C-2
24	40	100.0	467	4	US-08-496-841C-4
25	40	100.0	467	4	US-08-496-841C-134
26	40	100.0	467	4	US-08-496-841C-136
27	40	100.0	467	4	US-08-706-344C-2

28	40	100.0	467	4	US-08-706-344C-4	Sequence 4, Appl1
29	40	100.0	467	4	US-08-706-344C-28	Sequence 28, Appl1
30	40	100.0	467	4	US-08-706-344C-30	Sequence 30, Appl1
31	40	100.0	467	4	US-08-706-344C-32	Sequence 32, Appl1
32	30	75.0	679	4	US-08-214-583-2	Sequence 2, Appl1
33	29	72.5	348	4	US-09-248-528-7	Sequence 7, Appl1
34	29	72.5	348	4	US-09-549-108-7	Sequence 7, Appl1
35	29	72.5	348	4	US-09-549-111-7	Sequence 7, Appl1
36	29	72.5	348	4	US-09-549-111-7	Sequence 7, Appl1
37	28	70.0	250	1	US-08-378-761A-78	Sequence 78, Appl1
38	28	70.0	250	1	US-08-485-286-78	Sequence 78, Appl1
39	28	70.0	263	1	US-07-901-707-7	Sequence 7, Appl1
40	28	70.0	263	1	US-07-988-430-7	Sequence 7, Appl1
41	28	70.0	263	1	US-08-425-336-7	Sequence 7, Appl1
42	28	70.0	263	1	US-08-488-113B-7	Sequence 7, Appl1
43	28	70.0	263	1	US-08-477-484B-7	Sequence 7, Appl1
44	28	70.0	263	2	US-08-646-360-7	Sequence 7, Appl1
45	28	70.0	263	5	PCT-US92-09487-7	Sequence 7, Appl1

ALIGNMENTS

RESULT 1
US-08-875-972-4
; Sequence 4, Application US/08875972
; Patent No. 5985564
; GENERAL INFORMATION:
; APPLICANT: Huntington Potter and Jinhue Li
; TITLE OF INVENTION: ASSAY FOR IDENTIFYING GENES CAUSING
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173-4799
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/875, 972
; FILING DATE: 08-AUG-97
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/002,448
; FILING DATE: 16-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan Esq., Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: H095-03PA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781) 861-6240
; TELEFAX: (781) 861-9540
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 407 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-875-972-4

Query Match 100.0%; Score 40; DB 2; Length 407;
Best local Similarity 100.0%; Pred. No. 0.39;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LVNNAEGD 8
|||||||

Db 235 LVNARECD 242

RESULT 2
US-08-788-231A-15
Patent No. 597313
GENERAL INFORMATION:
APPLICANT: L'Hernault, Steven W.
TITLE OF INVENTION: SPE-4 RELATED PEPTIDES, ANTIBODIES AND
METHODS OF USE THEREOF
INVENTOR: L'Hernault, Steven W.
CORRESPONDENCE ADDRESS:
ADDRESS: Greenlee, Winner and Sullivan, P.C.
CITY: Boulder
STATE: Colorado
ZIP: 80303
COUNTRY: US
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: FASTED, Version 1.5
CURRENT APPLICATION DATA: Case #1.0, Version #1.30
APPLICATION NUMBER: US/08/788,231A
FILING DATE: 24-JAN-1997
CLASSIFICATION: 424
PRIORITY INFORMATION DATA:
PRIORITY NUMBER: US 60/010,672
FILING DATE: 126-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Feiber, Donna M.
REGISTRATION NUMBER: 33,878
TELEPHONE: (303) 499-8089
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 499-8089
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: not relevant
MOLECULE TYPE: protein
HYPOTHEICAL: YES
US-08-788-231A-15

Query Match 100.0%; Score 40; DB 3; Length 462;
Best Local Similarity 100.0%; Pred. No. 0.46;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LVNARECD 8
Db 290 LVNARECD 297

RESULT 3
US-08-670-479-18
Sequence 18, Application US/08670479
Patent No. 597313
GENERAL INFORMATION:
APPLICANT: Hardy, John A.
TITLE OF INVENTION: ALZHEIMER'S DISEASE
AND GENETICALLY-ASSOCIATED GENES
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESS: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
STATE: King of Prussia
COUNTRY: U.S.A.

ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTED, Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/670,479
FILING DATE: 26-JUN-1996
CLASSIFICATION: 435
PRIORITY INFORMATION DATA:
PRIORITY NUMBER: 60/001,500
FILING DATE: 18-JUL-1995
APPLICATION NUMBER: 60/001,800
FILING DATE: 02-AUG-1995
ATTORNEY/AGENT INFORMATION:
NAME: Han, William T.
REGISTRATION NUMBER: 34,344
TELEPHONE: 610-270-5219
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5090
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 463 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Peptide
HYPOTHEICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
US-08-670-479-18

Query Match 100.0%; Score 40; DB 2; Length 463;
Best Local Similarity 100.0%; Pred. No. 0.46;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LVNARECD 8
Db 291 LVNARECD 298

RESULT 4
US-08-670-964-4
Sequence 4, Application US/08670964
Patent No. 6010874
GENERAL INFORMATION:
APPLICANT: Hardy, John A.
TITLE OF INVENTION: ALZHEIMER'S DISEASE
AND GENETICALLY-ASSOCIATED GENES
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: SmithKline Beecham Corporation
STREET: 709 Swedeland Road - UM2220; P.O. Box 15
STATE: Philadelphia
COUNTRY: USA
ZIP: 19406
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTED, Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/670,964
FILING DATE: 26-JUN-1996
CLASSIFICATION: 435
PRIORITY INFORMATION DATA:
PRIORITY NUMBER: 60/001,142

FILING DATE: 13-JUL-1995
APPLICATION NUMBER: 60/001,501
FILING DATE: 18-JUL-1995
ATTORNEY/AGENT INFORMATION:
NAME: Han, William T
REGISTRATION NUMBER: 34,344
REFERENCE/DOCKET NUMBER: P50358
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5219
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 463 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-670-964-4

Query Match 100.0%; Score 40; DB 3; Length 463;
Best Local Similarity 100.0%; Pred. NO. 0.46;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LVNMAEGD 8
|||||
DB 291 LVNMAEGD 298

RESULT 5
US-08-888-077A-4
Sequence 4, Application US/08888077A
Patent No. 6020143
GENERAL INFORMATION:
APPLICANT: ST. GEORGE-HYSLOP, PETER H
APPLICANT: ROMMENS, JOHANNA M
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
TITLE OF INVENTION: TO ALZHEIMER'S DISEASE AND USUS THEREFOR.
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: LERNER, DAVID, LITTENBERG, KRUMHOLZ & MENTLIK
STREET: 600 SOUTH AVENUE WEST
CITY: WESTFIELD
STATE: NJ
COUNTRY: USA
ZIP: 07090-1497
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/888,077A
FILING DATE: 03-JUL-1997
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/592,541
FILING DATE: 26-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: PALISI, THOMAS M
REGISTRATION NUMBER: 36,629
REFERENCE/DOCKET NUMBER: SCHERING 3.0-017 CIP CIP IV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 654-5000
TELEFAX: (908) 654-7866
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 463 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-888-077A-4

Query Match 100.0%; Score 40; DB 3; Length 463;
Best Local Similarity 100.0%; Pred. NO. 0.46;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LVNMAEGD 8
|||||
DB 291 LVNMAEGD 298

RESULT 6
US-08-788-231A-17
Sequence 17, Application US/08788231A
Patent No. 6019974
GENERAL INFORMATION:
APPLICANT: L'Hernault, Steven W.
TITLE OF INVENTION: SPE-4 RELATED PEPTIDES, ANTIBODIES AND
TITLE OF INVENTION: METHODS
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: US
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/788,231A
FILING DATE: 24-JAN-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/010,672
FILING DATE: 26-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Feider, Donna M.
REGISTRATION NUMBER: 33,878
REFERENCE/DOCKET NUMBER: 60-95
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 499-8080
TELEFAX: (303) 499-8089
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 465 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: not relevant
MOLECULE TYPE: protein
HYPOTHETICAL: YES
US-08-788-231A-17

Query Match 100.0%; Score 40; DB 3; Length 465;
Best Local Similarity 100.0%; Pred. NO. 0.46;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LVNMAEGD 8
|||||
DB 293 LVNMAEGD 300

RESULT 7
US-08-967-101-2
Sequence 2, Application US/08967101
Patent No. 5840540
GENERAL INFORMATION:
APPLICANT: ST. GEORGE-HYSLOP, PETER H

APPLICANT: ROMMENS, JOHANNA M
APPLICANT: FRASER, PAUL E
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
TO ALZHEIMER'S DISEASE
NUMBER OF SEQUENCES: 183
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: High Street Tower - 125 High Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/967,101
FILING DATE: 10-NOV-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/592,541
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pitcher, Edmund R.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 467 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-967-101-2

Query Match 100.0%; Score 40; DB 2; Length 467;
Best Local Similarity 100.0%; Pred. No. 0.46;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVNMAEGD 8
Db 295 LVNMAEGD 302

RESULT 8
US-08-967-101-4
Sequence 4, Application US/08967101
Patent No. 5840540
GENERAL INFORMATION:
APPLICANT: ST. GEORGE-HYSLOP, PETER H
APPLICANT: ROMMENS, JOHANNA M
APPLICANT: FRASER, PAUL E
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
TO ALZHEIMER'S DISEASE
NUMBER OF SEQUENCES: 183
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: High Street Tower - 125 High Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/967,101

FILING DATE: 10-NOV-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/592,541
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pitcher, Edmund R.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 467 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-967-101-4

Query Match 100.0%; Score 40; DB 2; Length 467;
Best Local Similarity 100.0%; Pred. No. 0.46;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVNMAEGD 8
Db 295 LVNMAEGD 302

RESULT 9
US-08-967-101-134
Sequence 134, Application US/08967101
Patent No. 5840540
GENERAL INFORMATION:
APPLICANT: ST. GEORGE-HYSLOP, PETER H
APPLICANT: ROMMENS, JOHANNA M
APPLICANT: FRASER, PAUL E
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
TO ALZHEIMER'S DISEASE
NUMBER OF SEQUENCES: 183
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: High Street Tower - 125 High Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/967,101
FILING DATE: 10-NOV-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/592,541
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pitcher, Edmund R.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 134:
SEQUENCE CHARACTERISTICS:
LENGTH: 467 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-967-101-134

Query Match 100.0%; Score 40; DB 2; Length 467;
Best Local Similarity 100.0%; Pred. No. 0.46;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVNMAEGD 8
|||||
DB 295 LVNMAEGD 302

RESULT 10
US-08-592-541-2
; Sequence 2, Application US/08592541
; Patent No. 5986054
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
; APPLICANT: ROMMENS, JOHANNA M
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
; TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 183
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: High Street Tower - 125 High Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,541
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Pitcher, Edmund R.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 467 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-592-541-2

Query Match 100.0%; Score 40; DB 2; Length 467;
Best Local Similarity 100.0%; Pred. No. 0.46;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVNMAEGD 8
|||||
DB 295 LVNMAEGD 302

RESULT 11
US-08-592-541-4
; Sequence 4, Application US/08592541
; Patent No. 5986054
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
; APPLICANT: ROMMENS, JOHANNA M
; APPLICANT: FRASER, PAUL E
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
; TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 183
; CORRESPONDENCE ADDRESS:

ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: High Street Tower - 125 High Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,541
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Pitcher, Edmund R.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 467 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-592-541-4

Query Match 100.0%; Score 40; DB 2; Length 467;
Best Local Similarity 100.0%; Pred. No. 0.46;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVNMAEGD 8
|||||
DB 295 LVNMAEGD 302

RESULT 12
US-08-592-541-134
; Sequence 134, Application US/08592541
; Patent No. 5986054
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
; APPLICANT: ROMMENS, JOHANNA M
; APPLICANT: FRASER, PAUL E
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
; TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 183
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: High Street Tower - 125 High Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,541
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Pitcher, Edmund R.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 134:
; SEQUENCE CHARACTERISTICS:

LENGTH: 467 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-592-541-134

Query Match 100.0% Score 40; DB 2; Length 467;
Best Local Similarity 100.0% Pred. No. 0.46; 0;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVNMAEGD 8
DB 295 LVNMAEGD 302

RESULT 13
US-08-923-454A-10
Sequence 10, Application US/08923454A
Patent No. 6004794
GENERAL INFORMATION:
APPLICANT: Creasy, Caretha
APPLICANT: Layl, George
APPLICANT: Karian, Eric
APPLICANT: Clinkenbeard, Helen
APPLICANT: Browne, Michael
APPLICANT: Southan, Christopher
TITLE OF INVENTION: HUMAN SERINE PROTEASE
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/923,454A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/025436
FILING DATE: 06-SEPT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Baumeister, Kirk
REGISTRATION NUMBER: 33,833
REFERENCE/DOCKET NUMBER: P50547
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5096
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 467 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
US-08-923-454A-10

Query Match 100.0% Score 40; DB 3; Length 467;
Best Local Similarity 100.0% Pred. No. 0.46;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LVNMAEGD 8
DB 295 LVNMAEGD 302

RESULT 14
US-08-670-964-2
Sequence 2, Application US/08670964
Patent No. 6010874
GENERAL INFORMATION:
APPLICANT: Hardy, John A.
TITLE OF INVENTION: EARLY ONSET ALZHEIMER'S DISEASE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road - UW2220; P.O. Box 15
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19406
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/670,964
FILING DATE: 26-JUN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/001,142
FILING DATE: 13-JUL-1995
APPLICATION NUMBER: 60/001,501
FILING DATE: 18-JUL-1995
ATTORNEY/AGENT INFORMATION:
NAME: Han, William T
REGISTRATION NUMBER: 34,344
REFERENCE/DOCKET NUMBER: P50358
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5219
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 467 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-670-964-2

Query Match 100.0% Score 40; DB 3; Length 467;
Best Local Similarity 100.0% Pred. No. 0.46;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

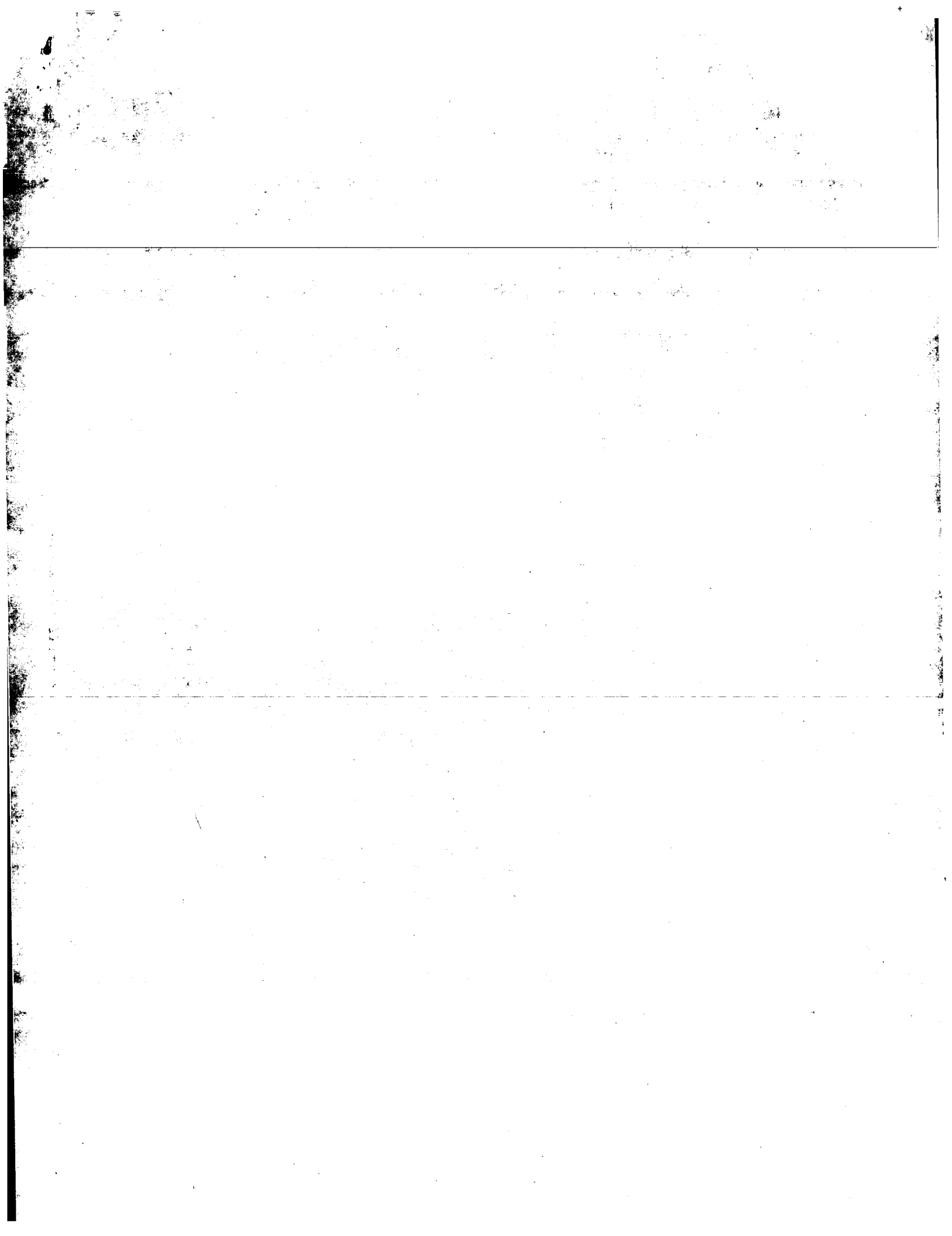
QY 1 LVNMAEGD 8
DB 295 LVNMAEGD 302

RESULT 15
US-08-888-077A-2
Sequence 2, Application US/08888077A
Patent No. 6020143
GENERAL INFORMATION:
APPLICANT: ST. GEORGE-HYSLOP, PETER H
APPLICANT: ROMMENS, JOHANNA M
APPLICANT: FRASER, PAUL E
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED

TITLE OF INVENTION: TO ALZHEIMER'S DISEASE AND USES THEREFOR.
 :
 : NUMBER OF SEQUENCES: 41
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: LERNER, DAVID, LITTENBERG, KRUMHOLTZ & MENTLIK
 : STREET: 600 SOUTH AVENUE WEST
 : CITY: WESTFIELD
 : STATE: NJ
 : COUNTRY: USA
 : ZIP: 07090-1497
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: ASCII
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/888,077A
 : FILING DATE: 03-JUL-1997
 : CLASSIFICATION: 530
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/592,541
 : FILING DATE: 26-JAN-1996
 : ATTORNEY/AGENT INFORMATION:
 : NAME: PALISI, THOMAS M
 : REGISTRATION NUMBER: 36,629
 : REFERENCE/DOCKET NUMBER: SCHERING 3.0-017 CIP CIP IV
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (908) 654-5000
 : TELEFAX: (908) 654-7866
 : INFORMATION FOR SEQ ID NO: 2:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 467 amino acids
 : TYPE: amino acid
 : TOPOLOGY: linear
 : MOLECULE TYPE: protein
 :
 : US-08-888-077A-2

Query Match 100.0%; Score 40; DB 3; Length 467;
 Best Local Similarity 100.0%; Pred. No. 0.46;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LVNMAEGD 8
 |||||||
 Db 295 LVNMAEGD 302

Search completed: September 6, 2001, 16:39:37
 Job time: 126 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:43:29 ; Search time 225.25 Seconds
(without alignments)
2.153 Million cell updates/sec

Title: US-09-603-713-9
Perfect score: 40
Sequence: 1 LVNMEGD 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_0601.*
1: /SID8/gcgdata/geneseq/AA1980.DAT:*
2: /SID8/gcgdata/geneseq/AA1981.DAT:*
3: /SID8/gcgdata/geneseq/AA1982.DAT:*
4: /SID8/gcgdata/geneseq/AA1983.DAT:*
5: /SID8/gcgdata/geneseq/AA1984.DAT:*
6: /SID8/gcgdata/geneseq/AA1985.DAT:*
7: /SID8/gcgdata/geneseq/AA1986.DAT:*
8: /SID8/gcgdata/geneseq/AA1987.DAT:*
9: /SID8/gcgdata/geneseq/AA1988.DAT:*
10: /SID8/gcgdata/geneseq/AA1989.DAT:*
11: /SID8/gcgdata/geneseq/AA1990.DAT:*
12: /SID8/gcgdata/geneseq/AA1991.DAT:*
13: /SID8/gcgdata/geneseq/AA1992.DAT:*
14: /SID8/gcgdata/geneseq/AA1993.DAT:*
15: /SID8/gcgdata/geneseq/AA1994.DAT:*
16: /SID8/gcgdata/geneseq/AA1995.DAT:*
17: /SID8/gcgdata/geneseq/AA1996.DAT:*
18: /SID8/gcgdata/geneseq/AA1997.DAT:*
19: /SID8/gcgdata/geneseq/AA1998.DAT:*
20: /SID8/gcgdata/geneseq/AA1999.DAT:*
21: /SID8/gcgdata/geneseq/AA2000.DAT:*
22: /SID8/gcgdata/geneseq/AA2001.DAT:*

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	100.0	8	22	AAB66579
2	40	100.0	8	22	AAB65341
3	40	100.0	101	21	AAB53687
4	40	100.0	125	18	AA11782
5	40	100.0	164	18	AA11787
6	40	100.0	407	18	AAW28507
7	40	100.0	429	19	AAW41429
8	40	100.0	430	17	AAW05761
9	40	100.0	434	17	AAW05760
10	40	100.0	463	17	AAW05734
11	40	100.0	463	18	AAW22948

12	40	100.0	463	18	AAW12376	Human S182 gene pr
13	40	100.0	463	18	AAW11840	Early onset Alzhei
14	40	100.0	463	19	AAV20854	Human presenilin I
15	40	100.0	463	19	AAW23965	Human presenilin I
16	40	100.0	463	19	AAW42375	Human presenilin I
17	40	100.0	465	21	AAV51393	Human S182 protein
18	40	100.0	465	21	AAV51394	Murine S182 protei
19	40	100.0	467	17	AAW05750	Presenilin-1-1 C26
20	40	100.0	467	17	AAW05751	Presenilin-1-1 P26
21	40	100.0	467	17	AAW05752	Presenilin-1-1 E28
22	40	100.0	467	17	AAW05753	Presenilin-1-1 A28
23	40	100.0	467	17	AAW05754	Presenilin-1-1 L28
24	40	100.0	467	17	AAW05755	Presenilin-1-1 G38
25	40	100.0	467	17	AAW05757	Presenilin-1-1 L39
26	40	100.0	467	17	AAW05758	Presenilin-1-1 C41
27	40	100.0	467	17	AAW05759	Presenilin-1-1 H
28	40	100.0	467	17	AAW05733	Murine presenilin.
29	40	100.0	467	17	AAW05735	Presenilin-1-1 A79
30	40	100.0	467	17	AAW05736	Presenilin-1-1 V82
31	40	100.0	467	17	AAW05737	Presenilin-1-1 V96
32	40	100.0	467	17	AAW05738	Presenilin-1-1 Y11
33	40	100.0	467	17	AAW05739	Presenilin-1-1 M13
34	40	100.0	467	17	AAW05740	Presenilin-1-1 I14
35	40	100.0	467	17	AAW05741	Presenilin-1-1 M14
36	40	100.0	467	17	AAW05742	Presenilin-1-1 H16
37	40	100.0	467	17	AAW05743	Presenilin-1-1 L17
38	40	100.0	467	17	AAW05744	Presenilin-1-1 G20
39	40	100.0	467	17	AAW05745	Presenilin-1-1 I21
40	40	100.0	467	17	AAW05746	Presenilin-1-1 I23
41	40	100.0	467	17	AAW05747	Presenilin-1-1 A24
42	40	100.0	467	17	AAW05748	Presenilin-1-1 A26
43	40	100.0	467	17	AAW05749	Human mutant S182
44	40	100.0	467	18	AAW27177	
45	40	100.0	467	18	AAW27176	Human S182 gene, P

ALIGNMENTS

RESULT 1	
ID AAB66579	standard. Peptide: 8 AA.
XX	
AC AAB66579;	
XX	
DT 12-APR-2001	(first entry)
XX	
DE Synthetic peptide PSI-gamma.	
XX	
KW Human; memapsin 2; neurotrophic; neuroprotective; amyloid precursor protein;	
KW APP; memapsin 2 inhibitor; Alzheimer's disease; PSI-gamma.	
XX	
OS Homo sapiens.	
OS Synthetic.	
PN WO200100665-A2.	
XX	
PD 04-JAN-2001.	
XX	
PF 27-JUN-2000;	2000WO-US17742.
XX	
PR 28-JUN-1999;	99US-0141363.
XX	
PR 30-NOV-1999;	99US-0168060.
XX	
PR 25-JAN-2000;	2000US-0177836.
XX	
PR 27-JAN-2000;	2000US-0178368.
XX	
PR 08-JUN-2000;	2000US-0210292.
XX	
PA (OKLA-) OKLAHOMA MEDICAL RES FOUND.	
XX	
PA (UNITI) UNIV ILLINOIS FOUND.	
XX	
PI Tang JUN, Hong L, Ghosh AK;	
XX	
DR WPI; 2001-137933/14.	

XX Novel memapsin 2 inhibitors which bind to active site of memapsin 2
 PT having 2 catalytic aspartic residues and substrate binding cleft, used
 PT to treat Alzheimer's disease by blocking amyloid precursor protein
 PT cleavage -
 XX
 PS Example 4; Page 33; 86pp; English.
 XX
 CC The present sequence is given in a specification relating to an inhibitor
 CC of catalytically active memapsin 2. The inhibitor binds to the memapsin 2
 CC active site, which is defined by the presence of two catalytic aspartic
 CC residues and a substrate binding cleft. The inhibitor is useful for
 CC the treatment and diagnosis of Alzheimer's disease. It is useful in
 CC screens for individuals with a genetic predisposition to Alzheimer's
 CC disease. The inhibitor is useful as a reagent for specifically binding to
 CC memapsin 2 or memapsin 2 analogues and for aiding in memapsin 2
 CC isolation, purification and characterisation.
 CC
 SQ Sequence 8 AA;

Query Match 100.0%; Score 40; DB 22; Length 8;
 Best Local Similarity 100.0%; Pred. No. 3.4e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LVNMAEGD 8
 Db 1 lvmaegd 8

RESULT 2

AA61341 ID AAB61341 standard; peptide; 8 AA.

AC AAB61341;

DT 02-APR-2001 (first entry)

DE Recombinant human memapsin 2 substrate.

XX Memapsin 2; catalyst; Alzheimer's.

OS Unidentified.

PN WO200100663-A2.

PD 04-JAN-2001.

PF 27-JUN-2000; 2000WO-US17661.

PR 28-JUN-1999; 99US-0141363.

PR 30-NOV-1999; 99US-0168060.

PR 25-JAN-2000; 2000US-0177836.

PR 27-JAN-2000; 2000US-0178368.

PR 08-JUN-2000; 2000US-0210292.

XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.

PI Tang JUN, Lin X, Koelsch G;

DR WPI; 2001-102885/11.

PT Purified recombinant catalytically active memapsin 2, used to screen
 PT inhibitors of it, which are used to treat and prevent Alzheimer's
 PT disease -

XX Disclosure; Page 24; 86pp; English.
 CC The present invention relates to a purified recombinant
 CC catalytically active memapsin 2. The invention may be used for
 CC isolating inhibitors which are used to treat or prevent
 CC Alzheimer's disease. The invention may also be used to screen
 CC for individuals more genetically prone to develop Alzheimer's
 CC disease.

CC disease.
 XX
 SQ Sequence 8 AA;

Query Match 100.0%; Score 40; DB 22; Length 8;
 Best Local Similarity 100.0%; Pred. No. 3.4e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LVNMAEGD 8
 Db 1 lvmaegd 8

RESULT 3

AA61367 ID AAB53687 standard; Protein; 101 AA.

AC AAB53687;

DT 09-MAR-2001 (first entry)

DE Human colon cancer antigen protein sequence SEQ ID NO:1227.

XX Human colon cancer; colon cancer antigen; diagnosis; detection;
 KW identification; cytostatic; cardioactive; neuroprotective; vulnary;
 KW immunomodulatory; muscular; gynaecological; gastrointestinal;
 KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;
 KW neural disorder; immune system disorder; muscular disorder;
 KW reproductive disorder; gastrointestinal disorder; renal disorder;
 KW infectious disease; cardiovascular disorder.

OS Homo sapiens.

PN WO200055351-A1.

PD 21-SEP-2000.

PF 08-MAR-2000; 2000WO-US05883.

PR 12-MAR-1999; 99US-0124270.

PR (HUMA-) HUMAN GENOME SCI INC.

PA Rosen CA, Ruben SM;

PI WPI; 2000-587534/55.

DR N-PSDB; AAC98444.

XX Colon cancer associated gene sequences, referred to as colon cancer
 XX antigens, useful for the treatment, prevention, and diagnosis of colon
 XX disorders such as colon cancer -

PS Claim 11; Page 1815-1816; 2104pp; English.

XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,
 CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
 CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
 CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
 CC vulnary, nephrotropic, antiinfective, antibacterial activities, and
 CC can be used in gene therapy. The colon cancer antigen polynucleotides,
 CC proteins and antibodies to the proteins are useful for the prevention,
 CC treatment and diagnosis of colon disorders, such as colon cancer. The
 CC polynucleotides may be used in diagnostics and research, such as for
 CC chromosome identification, and as hybridisation probes. The proteins
 CC may also be used to prevent diseases such as neural disorders, immune
 CC system disorders, muscular disorders, reproductive disorders, immune
 CC gastrointestinal disorders, wounds, renal disorders, infectious
 CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
 CC AAB54007 represent sequences used in the exemplification of the present
 CC invention.

SQ Sequence 101 AA;

Query Match 100.0%; Score 40; DB 21; Length 101;
 Best Local Similarity 100.0%; Pred. No. 0.24; Mismatches 8; Conservative 0; Indels 0; Gaps 0;

OY 1 LVNMAEGD 8
 |||||
 Db 49 lvnmaegd 56

RESULT 4
 AAM11782
 ID AAM11782 standard; Protein: 125 AA.

AC AAM11782;
 XX 24-APR-1997 (first entry)
 XX Early onset Alzheimer's disease gene product (T-cell lymphoma).
 DE Early onset Alzheimer's disease; EOAD: diagnosis; therapy;
 KW Early onset Alzheimer's disease; EOAD: diagnosis; therapy;
 KW expressed sequence tag; EST: T-cell lymphoma.
 KM Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 1 /note- "amino acid residue at position 1 is
 FT unidentified owing to degeneracy of the
 FT nucleotide sequence"

FT Misc-difference 6 /note- "amino acid residue at position 6 is
 FT unidentified owing to degeneracy of the
 FT nucleotide sequence"

FT Misc-difference 18 /note- "amino acid residue at position 18 is
 FT unidentified owing to degeneracy of the
 FT nucleotide sequence"

FT Misc-difference 21 /note- "amino acid residue at position 21 is
 FT unidentified owing to degeneracy of the
 FT nucleotide sequence"

FT Misc-difference 84 /note- "amino acid residue at position 84 is
 FT unidentified owing to degeneracy of the
 FT nucleotide sequence"

FT Misc-difference 90 /note- "amino acid residue at position 90 is
 FT unidentified owing to degeneracy of the
 FT nucleotide sequence"

FT Misc-difference 98 /note- "amino acid residue at position 98 is
 FT unidentified owing to degeneracy of the
 FT nucleotide sequence"

FT Misc-difference 113 /note- "amino acid residue at position 113 is
 FT unidentified owing to degeneracy of the
 FT nucleotide sequence"

FT Misc-difference 120 /note- "amino acid residue at position 120 is
 FT unidentified owing to degeneracy of the
 FT nucleotide sequence"

FT Misc-difference 122 /note- "amino acid residue at position 122 is
 FT unidentified owing to degeneracy of the
 FT nucleotide sequence"

PN W09701573-A2.

PD 16-JAN-1997.

PF 26-JUN-1996; 96WO-US11002.

XX 28-JUN-1995; 95US-0000590.
 PR (SMK) SMITHKLINE BEECHAM CORP.
 PA (SMK) SMITHKLINE BEECHAM PLC.
 XX

PI Karan EH;

DR WPI: 1997-100161/09.
 DR N-PSDB; AAT59472.

XX New Early Onset Alzheimer's Disease gene - used to develop prods.
 PT for diagnosis, detection of pre-disposition to, or treatment of
 PT Alzheimer's disease
 XX

PS Claim 3; Page 65-66; 97pp; English.

XX A set of polypeptides (AAM11780-85) have sequences deduced from the
 CC 3 reading frames, of both strands, of a human T-cell lymphoma cDNA
 CC clone (AAT59472) that corresponds to an expressed sequence tag (EST)
 CC of the early onset Alzheimer's disease (EOAD) gene. EOAD gene
 CC products (see also AAM11768-79 and AAM11786-97) can be expressed in
 CC host cells and used to screen for agonists or antagonists useful in
 CC EOAD therapy, or to raise antibodies useful in the diagnosis of
 CC EOAD or predisposition to EOAD.
 CC (N.B. in-frame stop codons in the 6 reading frames of the cDNA
 CC clone are ignored in the translated polypeptide sequences given in
 CC the specification)

XX Sequence 125 AA;

Query Match 100.0%; Score 40; DB 18; Length 125;
 Best Local Similarity 100.0%; Pred. No. 0.31; Mismatches 8; Conservative 0; Indels 0; Gaps 0;

OY 1 LVNMAEGD 8
 |||||
 Db 29 lvnmaegd 36

RESULT 5
 AAM11787
 ID AAM11787 standard; Protein: 164 AA.

XX AAM11787;

XX 24-APR-1997 (first entry)

XX Early onset Alzheimer's disease gene product (activated T-cell).

XX Early onset Alzheimer's disease; EOAD: diagnosis; therapy;
 KW expressed sequence tag; EST: activated T cell.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 144 /note- "amino acid residue at position 144 is
 FT unidentified owing to degeneracy of the
 FT nucleotide sequence"

FT Misc-difference 151 /note- "amino acid residue at position 151 is
 FT unidentified owing to degeneracy of the
 FT nucleotide sequence"

FT Misc-difference 160 /note- "amino acid residue at position 160 is
 FT unidentified owing to degeneracy of the
 FT nucleotide sequence"

PN W09701573-A2.

PD 16-JAN-1997.

```

XX PF 26-JUN-1996; 96WO-US11002.
XX XX
PR 28-JUN-1995; 95US-0000590.
XX XX
PA (SMIK ) SMITHKLINE BEECHAM CORP.
PA (SMIK ) SMITHKLINE BEECHAM PLC.
XX XX
PI Karan EH;
XX XX
DR WPI: 1997-100161/09.
DR N-PSDB: AAT59473.
XX XX
PT New Early Onset Alzheimer's Disease gene - used to develop prods.
PT for diagnosis, detection of pre-disposition to, or treatment of
PT Alzheimer's disease
XX XX
PS Claim 3; Page 69-70; 97pp; English.
XX XX
CC A set of polypeptides (AAW11786-91) have sequences deduced from the
CC 3 reading frames, of both strands, of a human activated T-cell cDNA
CC clone (AAT59473) that corresponds to an expressed sequence tag (EST)
CC of the early onset Alzheimer's disease (EOAD) gene. EOAD gene
CC products (see also AAW11768-85 and AAW11792-97) can be expressed in
CC host cells and used to screen for agonists or antagonists useful in
CC EOAD therapy, or to raise antibodies useful in the diagnosis of
CC EOAD or predisposition to EOAD.
CC (N.B. In-frame stop codons in the 6 reading frames of the cDNA
CC clone are ignored in the translated polypeptide sequences given in
CC the specification)
XX XX
SQ Sequence 164 AA;

Query Match 100.0%; Score 40; DB 18; Length 164;
Best Local Similarity 100.0%; Pred. No. 0.42;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LVNMAEGD 8
Db 4 lvnmaegd 11

RESULT 6
AAW28507
ID AAW28507 standard; Protein; 407 AA.
XX XX
AC AAW28507;
XX XX
DT 07-DEC-1997 (first entry)
XX XX
DE Partial AD3 sequence.
XX XX
AD3: AD4/AD3P; Alzheimer's disease; chromosome; missegregation;
KW presenilin; inhibitor; AD; trisomy 21.
XX XX
OS Homo sapiens.
XX XX
FH Key Location/Qualifiers
FH MISC-difference 86
FT /label= mutation
FT /note= "M -> L"
FT MISC-difference 103
FT /label= mutation
FT /note= "H -> R"
FT MISC-difference 186
FT /label= mutation
FT /note= "A -> E"
FT MISC-difference 226
FT /label= mutation
FT /note= "L -> V"
FT MISC-difference 350
FT /label= mutation

```

```

FT /note= "C -> Y"
XX XX
PN WO9707213-A2.
XX XX
PD 27-FEB-1997.
XX XX
PF 15-AUG-1996; 96WO-US13314.
XX XX
PR 16-AUG-1995; 95US-0002448.
XX XX
PA (HARD ) HARVARD COLLEGE.
XX XX
PI Li J, Potter H;
XX XX
DR WPI: 1997-165297/15.
DR N-PSDB: AAT87402.
XX XX
PT Identifying genes which cause chromosome missegregation - useful for
PT identifying causes of and treatments for diseases, e.g. Alzheimer's
PT disease, cancer and ageing
XX XX
PS Disclosure: Fig 1; 77pp; English.
XX XX
CC Identifying genes which cause improper chromosome segregation,
CC screening for inhibitors of chromosome missegregation and processes
CC caused by genes encoding chromosome missegregation promoters
CC was exemplified using Alzheimer's disease. The sequences
CC given in AAT87401 to AAT87426 can be used in the above methods.
CC The five mutations indicated in the Features Table cosegregate
CC with early-onset familial Alzheimer's disease. It is predicted
CC that these mutations result in increased levels of cells with
CC trisomy 21 in carriers of the mutation compared with non-carriers.
XX XX
SQ Sequence 407 AA;

Query Match 100.0%; Score 40; DB 18; Length 407;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LVNMAEGD 8
Db 235 lvnmaegd 242

RESULT 7
AAW41429
ID AAW41429 standard; Protein; 429 AA.
XX XX
AC AAW41429;
XX XX
DT 04-JUN-1998 (first entry)
XX XX
DE PS1/429 protein.
XX XX
KW Presenilin peptide; PS1/429; immunogen; immune response; PS1 gene;
KW Alzheimer's disease; mitochondrial pathology; neurodegeneration;
apoptosis.
XX XX
OS Homo sapiens.
XX XX
PN WO9746678-A1.
XX XX
PD 11-DEC-1997.
XX XX
PF 03-JUN-1997; 97WO-US09272.
XX XX
PR 18-JUL-1996; 96US-0683315.
XX XX
PR 06-JUN-1996; 96US-0659296.
XX XX
PA (FARB ) BAYER CORP.
XX XX
PI Chisholm JC, Davis JN, Drache B;

```


PA (HSCR-) HSC RES & DEV LP.
 PA (UTOR) UNIV TORONTO GOVERNING COUNCIL.
 XX Fraser PE, Rommens JM, St George-Hyslop PH;
 XX WPI: 1996-497631/49.
 DR
 XX
 PT New presentin genes - useful for diagnosis, therapy and drug
 PT screening of familial Alzheimer's disease, cerebral disorders, etc.
 XX
 PS Claim 3; Page -: 178pp; English.
 XX
 CC AAM05736-W05760 represent mutated versions of the human presentin-1-1
 CC protein (see AAM05733 for wild type sequence). AAM05734 represents a
 CC different wild type form of presentin-1 that results from alternate
 CC splicing of the genomic DNA sequence. The presentins are a family of
 CC highly conserved integral membrane proteins with a common structural
 CC motif, common alternate splicing patterns, and common mutational hot
 CC spot regions. Mutations in PS genes are implicated in familial
 CC Alzheimer's disease (AD) and possibly other diseases such as cerebral
 CC haemorrhage, schizophrenia, depression etc., so detection of mutations in
 CC the DNA encoding the wild type sequences can be used for diagnosis of
 CC these diseases. The wild type proteins, or vectors that express them or
 CC containing antisense sequences, antibodies selective for these mutant
 CC forms of the proteins and modulators of PS gene expression are
 CC potentially useful for treatment of AD etc. Transgenic animals are
 CC useful as models for drug screening. The antibodies can also be used e.g.
 CC for affinity purification and in immunoassays.
 CC
 XX Sequence 434 AA;
 SQ

Query Match 100.0%; Score 40; DB 17; Length 434;
 Best Local Similarity 100.0%; Pred. No. 1.3; Indels 0; Gaps 0;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LVNMAEGD 8
 Db 262 lvnmaegd 269

RESULT 10

AAM05734
 ID AAM05734 standard; Protein: 463 AA.

XX AAM05734;

XX 23-JUL-1997 (first entry)

XX Presentin-1-2.

XX Presentin-1; human: hPS1-1; hPS1-2; PS-2; integral membrane protein; AD;
 KW familial Alzheimer's disease; cerebral haemorrhage; schizophrenia;
 KW depression; antibody; gene expression modulator; therapy.

XX Homo sapiens.

XX W09634099-A2.

XX 31-OCT-1996.

XX 29-APR-1996; 96WO-CA00263.

XX 31-JUL-1995; 95US-0509359.

XX 28-APR-1995; 95US-0431048.

XX 28-JUN-1995; 95US-0496841.

XX (HSCR-) HSC RES & DEV LP.

XX (UTOR) UNIV TORONTO GOVERNING COUNCIL.

XX Fraser PE, Rommens JM, St George-Hyslop PH;
 XX WPI: 1996-497631/49.

DR N-PSDB; AAT40029.
 XX
 PT New presentin genes - useful for diagnosis, therapy and drug
 PT screening of familial Alzheimer's disease, cerebral disorders, etc.
 XX
 PS Claim 2; Page 132-134; 178pp; English.
 XX

CC AAM05733 and AAM05734 represent the two different forms of wild type
 CC human presentin-1 (PS-1). This form of presentin-1 results from
 CC alternate splicing of the genomic DNA sequence. AAM05762 represents the
 CC coding sequence for wild type human PS-2. The presentins are a family
 CC of highly conserved integral membrane proteins with a common structural
 CC motif, common alternate splicing patterns, and common mutational hot
 CC spot regions. Mutations in PS genes are implicated in familial
 CC Alzheimer's disease (AD) and possibly other diseases such as cerebral
 CC haemorrhage, schizophrenia, depression etc., so detection of mutations in
 CC the DNA encoding these sequences can be used for diagnosis of these
 CC diseases. These proteins, or vectors that express them or containing
 CC antisense sequences, antibodies selective for mutant forms of these
 CC proteins (such as AAM05736) and modulators of PS gene expression are
 CC potentially useful for treatment of AD etc. Transgenic animals are useful
 CC as models for drug screening. The antibodies can also be used e.g. for
 CC affinity purification and in immunoassays.
 CC
 XX Sequence 463 AA;
 SQ

Query Match 100.0%; Score 40; DB 17; Length 463;
 Best Local Similarity 100.0%; Pred. No. 1.4; Indels 0; Gaps 0;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LVNMAEGD 8
 Db 291 lvnmaegd 298

RESULT 11

AAM22948
 ID AAM22948 standard; Protein: 463 AA.

XX AAM22948;

XX 19-MAR-1998 (first entry)

XX Presentin-1 VRSQ variant.

XX Identification; determination; neurological disease susceptibility;
 KW detection; alternative splice site; polyadenylated mRNA transcript;
 KW familial Alzheimer's disease; FAD; presentin 1; VRSQ variant.

XX Homo sapiens.

XX EP791660-A1.

XX 27-AUG-1997.

XX 14-FEB-1997; 97EP-0300988.

XX 22-FEB-1996; 96US-0012077.

XX (SMIK) SMITHLINE BEECHAM CORP.

XX (USF-) UNIV SOUTH FLORIDA DEPT PSYCHIATRY.

XX (UNIW) UNIV WASHINGTON.

XX Barton A, Coate A, Hardy J;

XX WPI: 1997-418049/39.

XX N-PSDB; AAT75576.

XX Diagnosis of, or susceptibility to neurological disease -
 PT specifically Alzheimer's disease, by detecting aberrant splicing in
 PT mRNA
 XX

PS Example 1: Pages 9-11; 21pp; English.
 CC Identifying susceptibility to a neurological disease, comprises
 CC detecting an alternative splice site in a polyadenylated mRNA
 CC transcript in a sample of genetic material, where the alternative
 CC splice site encodes AAW22944, or detecting AAW22944 in the protein
 CC encoded by the mRNA. Tests on 3 early onset familial Alzheimer's
 CC disease (FAD) patients, 6 late onset sporadic Alzheimer's disease
 CC (AD) patients and 4 neurologically normal subjects, indicated that
 CC mRNA transcripts of the presenilin 1 gene in samples from various
 CC brain regions occur in 2 forms, PS-1-long (containing a VRSQ
 CC motif) and PS-1-short (lacking the VRSQ motif, i.e. the protein
 CC denoted by the present sequence), and that the PS-1-long levels
 CC in hippocampus and frontal cortex samples are significantly lower
 CC in FAD patients than in AD and normal subjects.
 CC
 SO Sequence 463 AA;
 Query Match 100.0%; Score 40; DB 18; Length 463;
 Best Local Similarity 100.0%; Pred. No. 1.4;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 LVNMAEGD 8
 |||||
 DB 291 lvnmaegd 298
 RESULT 12
 AAW12376
 ID AAW12376 standard; Protein: 463 AA.
 XX
 AC AAW12376;
 XX
 DT 17-JUN-1997 (first entry)
 XX
 DE Human S182 gene product.
 XX
 KW S182 gene; familial Alzheimer's disease; diagnosis;
 KW transgenic animal.
 XX
 OS Homo sapiens.
 XX
 PN WO9703999-A1.
 XX
 PD 06-FEB-1997.
 XX
 PF 26-JUN-1996; 96WO-US11065.
 XX
 PR 02-AUG-1995; 95US-0001800.
 PR 18-JUL-1995; 95US-0001500.
 XX
 PA (UYSF-) UNIV SOUTH FLORIDA.
 PA (UNIM) UNIV WASHINGTON SCHOOL MED.
 PI Goate AM, Hardy JA;
 XX
 DR MPI: 1997-132571/12.
 DR N-PSDB: AAT63207.
 XX
 PT New mutants of the S182 gene associated with familial Alzheimer's
 PT disease - and related protein and transgenic animals, useful as
 PT models for screening and assessing potential drugs
 XX
 PS Disclosure: Fig 1A-D; 26pp; English.
 XX
 CC A polypeptide (AAW12376) is the product of an S182 gene cDNA clone
 CC (AAT63207) isolated from a human brain library. Several mutations in
 CC the S182 gene have been found in families with members affected by
 CC early onset Alzheimer's disease (AD): in 2 families Met to Val at
 CC position 135; in 3 families Met to Val at 142; in 1 family Pro to
 CC Ser at 263; in 4 families Glu to Ala at 276; and in 1 family Glu to
 CC Gly at 280. Detection of the mutations is used to diagnose AD, or

CC a predisposition to it. Transgenic animals can be produced that
 CC are useful as models for screening and assessing potential drugs.
 CC
 SO Sequence 463 AA;
 Query Match 100.0%; Score 40; DB 18; Length 463;
 Best Local Similarity 100.0%; Pred. No. 1.4;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 LVNMAEGD 8
 |||||
 DB 291 lvnmaegd 298
 RESULT 13
 AAW11840
 ID AAW11840 standard; Protein: 463 AA.
 XX
 AC AAW11840;
 XX
 DT 07-MAY-1997 (first entry)
 XX
 DE Early onset Alzheimer's disease (EOAD) splice variant polypeptide.
 XX
 KW Early onset Alzheimer's disease; EOAD; neurodegenerative disease;
 KW diagnosis; therapy; inhibitor; antagonist; antibody.
 XX
 OS Homo sapiens.
 XX
 PN WO9703086-A1.
 XX
 PD 30-JAN-1997.
 XX
 PF 26-JUN-1996; 96WO-US11064.
 XX
 PR 18-JUL-1995; 95US-0001501.
 PR 13-JUL-1995; 95US-0001142.
 XX
 PA (UYSF-) UNIV SOUTH FLORIDA.
 PA
 PI Hardy JA;
 XX
 DR MPI: 1997-118980/11.
 DR N-PSDB: AAT59536.
 XX
 PT Early onset Alzheimer's disease gene - useful for diagnosing a
 PT pre-disposition to Alzheimer's disease
 XX
 PS Disclosure: Fig 2; 44pp; English.
 XX
 CC A 463-amino acid polypeptide (AAW11840) is the product of a full-
 CC length cDNA (AAT59536) of an early onset Alzheimer's disease (EOAD)
 CC splice variant gene. A 467-amino acid polypeptide (AAW11839) is the
 CC product of a full-length cDNA (AAT59535) of the EOAD gene. The 2
 CC polypeptides can be produced in transformed host cells and used to
 CC raise antibodies, or to identify antagonist/inhibitor cpds. useful
 CC in the treatment of Alzheimer's disease, esp. EOAD.
 XX
 SO Sequence 463 AA;
 Query Match 100.0%; Score 40; DB 18; Length 463;
 Best Local Similarity 100.0%; Pred. No. 1.4;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 LVNMAEGD 8
 |||||
 DB 291 lvnmaegd 298
 RESULT 14
 AAT20854

ID AAY20854 standard; Protein: 463 AA.
 AC AAY20854;
 DT 22-JUL-1999 (first entry)
 DE Human presentinlin I wild type protein fragment.
 KW Human; beta-amyloid precursor protein; beta-APP; diagnosis; cancer;
 KW frameshift mutation; age-related disease; neurodegenerative disorder;
 KW Alzheimer's disease; Down's syndrome; myotonic dystrophy; neuronal;
 KW Huntington's disease; multiple sclerosis; alcoholic liver disease;
 KW diabetes mellitus type II; microtubule associated protein; Tau; Big Tau;
 KW ubiquitin B; apolipoprotein E; MAP2; neurofilament-L; neurofilament-M;
 KW neurofilament-F; presentinlin I; presentinlin II; cellular tumour antigen;
 KW glial fibrillary acidic protein; GFAP; p53; semaphorin III; HUPF-1;
 KW bcl-2; B-cell leukemia/lymphoma 2 proto-oncogene; HMG-C; NSP-A;
 KW high mobility group protein-C; neuroendocrine specific protein A.
 OS Homo sapiens.
 PN WO9845322-A2.
 XX 15-OCT-1998.
 PD 02-APR-1998; 98WO-1B00705.
 PF 10-APR-1997; 97US-0043163.
 PR (UYUT-) RIKSUNIV UTRECHT.
 PA (ROYA-) ROYAL NETHERLANDS ACAD ARTS & SCI.
 PY (UYRO-) UNIV ROTTERDAM ERASMUS.
 PI Burbach JPH, Grosvelde FG, Van Leeuwen FW;
 DR WPI: 1998-609901/51.
 DR N-PSDB: AAX75761.
 XX Diagnosing disease by detecting frameshift mutations in RNA or
 PT corresponding protein mutations - used to diagnose cancer and
 PT neurological diseases, particularly Alzheimer's disease, and also
 PT for treatment and prevention with specific ribozymes or wild-type
 PT RNA
 PS Disclosure: Figure 10: 258pp: English.
 XX This invention describes a novel method for the diagnosis of a disease
 CC caused by, or associated with, an RNA molecule that has a frameshift
 CC mutation. The method is used to diagnose age-related diseases, especially
 CC cancer and a wide range of neurodegenerative disorders (e.g. Alzheimer's
 CC disease, Down's syndrome, myotonic dystrophy, Huntington's disease,
 CC multiple sclerosis, alcoholic liver disease, diabetes mellitus type II
 CC and many others listed) or susceptibility to these disorders. The method
 CC allows a definitive diagnosis of Alzheimer's disease in living patients,
 CC at an early stage. It is based on the observation that disease may be
 CC caused by mutations in RNA rather than DNA. The invention describes the
 CC use of neuronal system RNA molecules, specifically proteins including
 CC beta-amyloid precursor protein (beta-APP), the microtubule associated
 CC proteins Tau and Big Tau, ubiquitin B, apolipoprotein E, microtubule
 CC associated protein 2 (MAP2), neurofilament-L, neurofilament-M,
 CC neurofilament-F, presentinlin I, presentinlin II, glial fibrillary acidic
 CC protein (GFAP), the cellular tumour antigen p53, B-cell leukemia/lymphoma
 CC 2 (bcl-2) proto-oncogene, semaphorin III, HUPF-1, high mobility group
 CC protein-C (HMG-C) and neuroendocrine specific protein A.
 XX Sequence 463 AA:
 SQ

Query Match 100.0%; Score 40; DB 19; Length 463;
 Best Local Similarity 100.0%; Pred. No. 1.4;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 LVNMAEGD 8

Db 291 lvnmaegd 298
 RESULT 15
 ID AAW23965 standard; Protein: 463 AA.
 AC AAW23965;
 DT 20-JUL-1998 (first entry)
 DE Human presentinlin-1.
 KW Presentinlin-1; p51 gene; human; familial Alzheimer's disease; FAD;
 KW cerebral haemorrhage; schizophrenia; depression; epilepsy;
 KW mental retardation; diagnosis; therapy; transgenic animal.
 OS Homo sapiens.
 PN WO9801549-A2.
 XX 15-JAN-1998.
 PD 04-JUL-1997; 97WO-CA00475.
 PF 02-JAN-1997; 97US-0034590.
 PR 05-JUL-1996; 96US-0021673.
 PR 12-JUL-1996; 96US-0021700.
 PR 08-NOV-1996; 96US-0028895.
 PA (HSCR-) HSC RES & DEV LP.
 PY (UTOR) UNIV TORONTO GOVERNING COUNCIL.
 PI Fraser PE, Rommens JM, St George-Hyslop PH;
 DR WPI: 1998-286355/25.
 DR N-PSDB: AAV04667.
 XX New isolated mutant presentinlin-1 genes - useful for developing
 PT products for use in detection, diagnosis and therapy of Alzheimer's
 PT disease and for drug screening
 PS Claim 1: Page 185-186; 238pp; English.
 XX This polypeptide comprises human presentinlin-1 (hps1). Its amino
 CC acid sequence was deduced from an isolated cDNA clone (see AAV04667).
 CC Another hps1 sequence (see AAW23964) results from alternative
 CC splicing of the hps1 mRNA transcript. A murine p51 homologue (see
 CC AAW23966) and a human presentinlin-2 protein (see AAW23967) are also
 CC provided. Mutations in the p5-1 and p5-2 genes are linked to the
 CC development in humans of forms of familial Alzheimer's disease
 CC (FAD) and may be causative of other disorders, e.g. cognitive,
 CC intellectual, neurological or physiological disorders such as
 CC cerebral haemorrhage, schizophrenia, depression, mental retardation
 CC and epilepsy. Use of the nucleic acids and proteins comprising or
 CC derived from the presentinlins is made in screening and diagnosing
 CC FAD, identifying and developing therapeutics for treatment of FAD,
 CC and in producing cell lines and transgenic animals useful as models
 CC of FAD. Methods for identifying substances that bind to, or
 CC modulate the activity of a presentinlin protein, and methods for

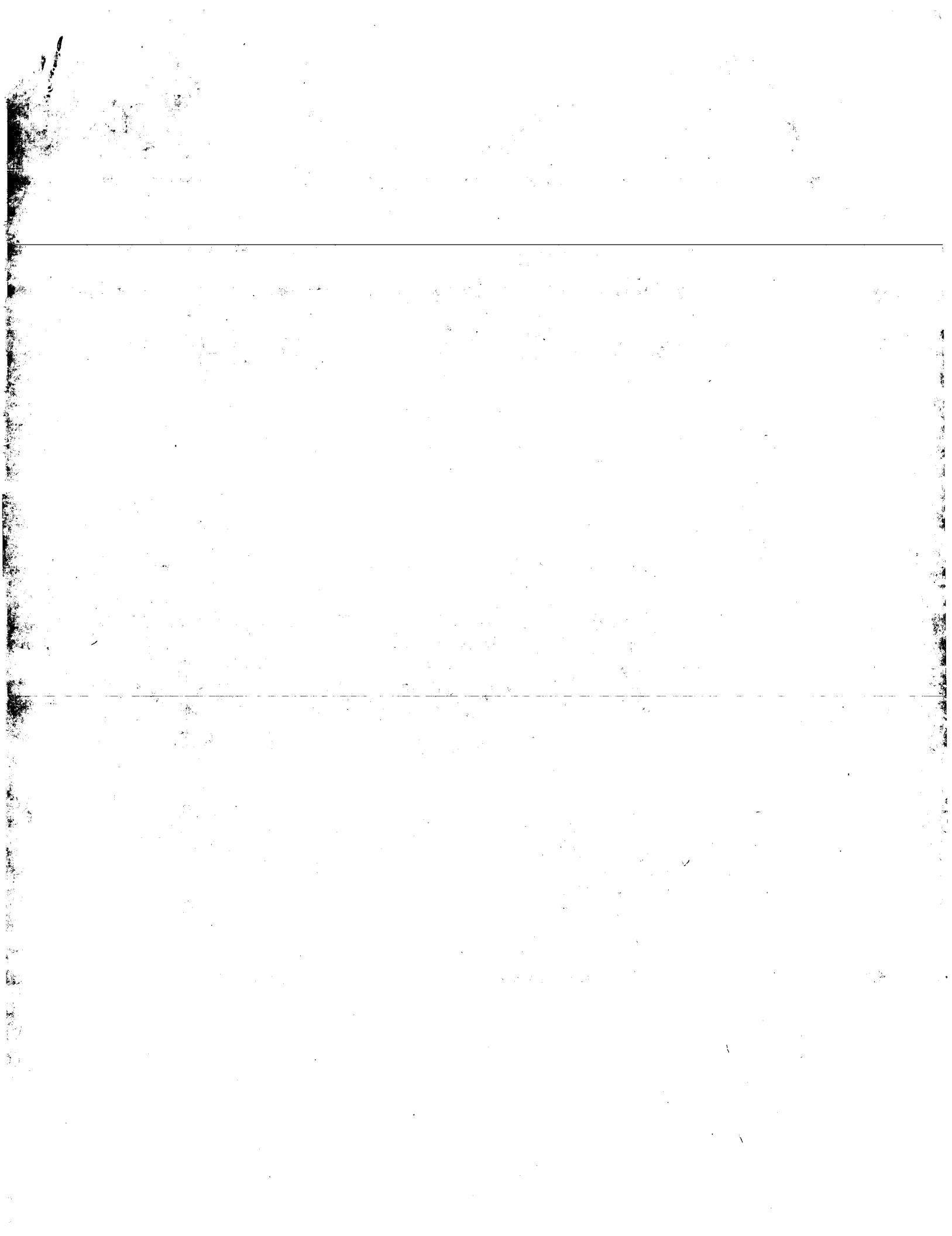
CC Identifying substances that affect the interaction of a
 CC presenilin-interacting protein with a presenilin protein are also
 CC disclosed.
 XX

Sequence 463 AA:

Query Match 100.0%; Score 40; DB 19; Length 463;
 Best Local Similarity 100.0%; Pred. No. 1.4;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LVNMAEGD 8
 |||||
 Db 291 lvnmaegd 298

Search completed: September 6, 2001, 16:43:29
 Job time: 358 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:49:48 ; Search time 231.42 Seconds
(without alignments)
2.287 Million cell updates/sec

Title: US-09-603-713-8
Perfect score: 21
Sequence: 1 DMSG 4

Scoring table: BL0SUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues
Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SPTRMBL_16:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_phage:*
11: sp_plant:*
12: sp_rodent:*
13: sp_unclassified:*
14: sp_vertebrate:*
15: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	100.0	13	4	Q9UPES5
2	21	100.0	15	6	Q9TRR9
3	21	100.0	19	6	Q9TRR4
4	21	100.0	20	6	Q9TRR4
5	21	100.0	24	10	Q9S937
6	21	100.0	25	2	Q69137
7	21	100.0	34	11	Q9OV52
8	21	100.0	35	8	Q9GF85
9	21	100.0	35	14	Q11563
10	21	100.0	36	6	Q9TSA3
11	21	100.0	41	6	Q9TSA3
12	21	100.0	45	6	Q9TSA3
13	21	100.0	48	14	Q83190
14	21	100.0	48	14	Q83195
15	21	100.0	48	14	Q83200
16	21	100.0	48	14	Q83205
17	21	100.0	48	14	Q83209
18	21	100.0	48	14	Q83214
19	21	100.0	48	14	Q83220

20	21	100.0	48	14	Q83225	Q83225 mouse cytom
21	21	100.0	48	14	Q83230	Q83230 mouse cytom
22	21	100.0	48	14	Q83233	Q83233 mouse cytom
23	21	100.0	48	14	Q83238	Q83238 mouse cytom
24	21	100.0	48	14	Q83244	Q83244 mouse cytom
25	21	100.0	48	14	Q83275	Q83275 mouse cytom
26	21	100.0	48	14	Q83282	Q83282 mouse cytom
27	21	100.0	52	2	Q9R5H1	Q9R5H1 mouse cytom
28	21	100.0	55	10	Q41920	Q41920 arabidopsis
29	21	100.0	57	5	Q9GWS3	Q9GWS3 leishmania
30	21	100.0	58	14	Q69128	Q69128 human herpe
31	21	100.0	60	10	Q41954	Q41954 arabidopsis
32	21	100.0	69	2	Q9R1J2	Q9R1J2 streptococ
33	21	100.0	74	2	Q31257	Q31257 anabaena sp
34	21	100.0	79	14	Q86982	Q86982 zucchini ye
35	21	100.0	80	2	Q9PD45	Q9PD45 xyella fas
36	21	100.0	82	2	Q87515	Q87515 escherichia
37	21	100.0	82	14	Q69144	Q69144 human herpe
38	21	100.0	85	10	Q22378	Q22378 glycine max
39	21	100.0	88	10	Q9XGV0	Q9XGV0 orobanche r
40	21	100.0	90	2	Q84334	Q84334 chlamydia t
41	21	100.0	91	14	Q97256	Q97256 human immu
42	21	100.0	93	10	P81224	P81224 oryza sativ
43	21	100.0	93	13	Q9PTF8	Q9PTF8 brachydanio
44	21	100.0	93	14	Q41098	Q41098 paramyctium
45	21	100.0	96	2	Q9HVT9	Q9HVT9 pseudomonas

ALIGNMENTS

RESULT 1
ID Q9UPES5 PRELIMINARY; PRT; 13 AA.
AC Q9UPES5:
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE INOSINE MONOPHOSPHATASE 2 (FRAGMENT).
GN IMPA2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-97463449; PubMed-9322233;
RA Yoshikawa T., Turner G., Esterling L.E., Sanders A.R.,
RA DeTeira-Wadleigh S.D.;
RT "A novel human myo-inositol monophosphatase gene, IMP.18p, maps to a
RT susceptibility region for bipolar disorder.*";
RL MOL. Psych. 2:393-397(1997).
DR EMBL; AF025866; AAD22138.1;
DR EMBL; AF025865; AAD22138.1; JOINED.
FT NON_TER 1
FT NON_TER 13
SQ SEQUENCE 13 AA; 1331 MW; 89C724C8E3457865 CRC64;

Query Match 100.0%; Score 21; DB 4; Length 13;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DMSG 4
Db 4 DMSG 7
RESULT 2
ID Q9TRR9 PRELIMINARY; PRT; 15 AA.
AC Q9TRR9:
DT 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE CALCYCLIN-ASSOCIATED PROTEIN, CAP50-CA2+/PHOSPHOLIPID-BINDING PROTEIN
 DE L-8 FRAGMENT.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=92250478; PubMed=1533622;
 RA Tokumitsu H., Mizutani A., Minami H., Kobayashi R., Hidaka H.;
 RT "A calyculin-associated protein is a newly identified member of the
 RT Ca2+/phospholipid-binding proteins, annexin family.",
 RL J. Biol. Chem. 267:8919-8924(1992).
 DR HSSP: P13214; IANN.
 SQ SEQUENCE 15 AA; 1686 MW; 299E23A743829077 CRC64;

Query Match 100.0%; Score 21; DB 6; Length 15;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTSG 4
 ||||
 Db 9 DTSG 12

RESULT 3
 ID 09TRR4 PRELIMINARY; PRT; 19 AA.
 AC 09TRR4;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE CALCYCLIN-ASSOCIATED PROTEIN, CAP50-CA2+/PHOSPHOLIPID-BINDING PROTEIN
 DE L-21 FRAGMENT.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=92250478; PubMed=1533622;
 RA Tokumitsu H., Mizutani A., Minami H., Kobayashi R., Hidaka H.;
 RT "A calyculin-associated protein is a newly identified member of the
 RT Ca2+/phospholipid-binding proteins, annexin family.",
 RL J. Biol. Chem. 267:8919-8924(1992).
 DR HSSP: P13134; IAVC.
 SQ SEQUENCE 19 AA; 2186 MW; 576DC1604E19BCB8 CRC64;

Query Match 100.0%; Score 21; DB 6; Length 19;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTSG 4
 ||||
 Db 9 DTSG 12

RESULT 4
 ID 09TR04 PRELIMINARY; PRT; 20 AA.
 AC 09TR04;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE CALCYCLIN-ASSOCIATED PROTEIN PEPTIDE L-9, CAP-50-ANNEXIN.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.

OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=92317074; PubMed=1618851;
 RA Mizutani A., Usuda N., Tokumitsu H., Minami H., Yasui K.,
 RA Kobayashi R., Hidaka H.;
 RT "CAP-50, a newly identified annexin, localizes in nuclei of cultured
 RT fibroblast 3Y1 cells."
 RL J. Biol. Chem. 267:13498-13504(1992).
 DR HSSP: P13214; IANN.
 SQ SEQUENCE 20 AA; 2296 MW; 89492C83166F4523 CRC64;

Query Match 100.0%; Score 21; DB 6; Length 20;
 Best Local Similarity 100.0%; Pred. No. 55;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTSG 4
 ||||
 Db 9 DTSG 12

RESULT 5
 ID 09S937 PRELIMINARY; PRT; 24 AA.
 AC 09S937;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
 DE H(+)-TRANSLOCATING (PYROPHOSPHATE-ENERGIZED) INORGANIC PYROPHOSPHATASE
 DE BETA-1 POLYPEPTIDE (EC 3.6.1.1) (FRAGMENT).
 OS Beta vulgaris (Sugar beet).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Caryophyllidae;
 OC Caryophyllales; Chenopodiaceae; Beta.
 OX NCBI_TaxID=3555;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=92179265; PubMed=1311852;
 RA Sarafian V., Kim Y., Poole R.J., Rea P.A.;
 RT "Molecular cloning and sequence of cDNA encoding the pyrophosphate-
 RT energized vacuolar membrane proton pump of Arabidopsis thaliana."
 RL Proc. Natl. Acad. Sci. U.S.A. 89:1775-1779(1992).
 SQ SEQUENCE 24 AA; 2396 MW; CE19F75ADBEPD43B CRC64;

Query Match 100.0%; Score 21; DB 10; Length 24;
 Best Local Similarity 100.0%; Pred. No. 67;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTSG 4
 ||||
 Db 14 DTSG 17

RESULT 6
 ID 069137 PRELIMINARY; PRT; 25 AA.
 AC 069137;
 DT 01-AUG-1998 (TREMBlrel. 07, Created)
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
 DT 01-AUG-1998 (TREMBlrel. 07, Last annotation update)
 DE HYPOTHETICAL 2.8 KDA PROTEIN (FRAGMENT).
 OS Rhodobacter capsulatus (Rhodospirillum rubrum).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
 OC Rhodobacter.
 OX NCBI_TaxID=1061;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SB 1003;
 RA Jiang Z., Bauer E.C.;
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF064095; AAC17426.1; -.

KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 25 AA; 2835 MW; 8CA72DB3C87ABAD4 CRC64;

Query Match 100.0%; Score 21; DB 2; Length 25;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DTSG 4
|||
Db 21 DTSG 24

RESULT 7
O9GV52 PRELIMINARY: PRT: 34 AA.
AC O9GV52.
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE ANNEXIN V.
OS Mus sp.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCB1_TaxID=10095;
RN [1]
RP SEQUENCE.
RX MEDLINE=94009914; PubMed=8405662;
RA Kristensen B.I., Kristensen P., Johansen A.H.;
RT "Identification of the major annexins in Ehrlich ascites tumor cells.";
RL Int. J. Biochem. 25:1195-1202(1993).
DR HSSP: P14668; IAB.
DR InterPro: IPR001464; -
DR SMART: SM00335; ANX; 1.
SQ SEQUENCE 34 AA; 3885 MW; 47DB3B9DE9E2B6A7 CRC64;

Query Match 100.0%; Score 21; DB 11; Length 34;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DTSG 4
|||
Db 27 DTSG 30

RESULT 8
O9GFB5 PRELIMINARY: PRT: 35 AA.
AC O9GFB5.
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE PHOTOSYSTEM II SUBUNIT (FRAGMENT).
GN PSBH.
OS Ginkgo biloba (Ginkgo).
OC Chloroplast.
CC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
CC Ginkgophyta; Ginkgoales; Ginkgoaceae; Ginkgo.
OX NCB1_TaxID=3311;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=11080123;
RA "Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms.";
RL Am. J. Bot. 87:1712-1730(2000).
DR EMBL: AF123851; AAC26284.1; -
KW Chloroplast.
FT NON_TER 35
SQ SEQUENCE 35 AA; 3688 MW; 770A544C1DE242D9 CRC64;

Query Match 100.0%; Score 21; DB 8; Length 35;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DTSG 4
|||
Db 8 DTSG 11

RESULT 9
O11563 PRELIMINARY: PRT: 35 AA.
AC O11563.
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-AUG-1998 (TREMBLrel. 07, Last annotation update)
DE ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCB1_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RA Balotta C., Riva C., Violin M., Colombo C., Galli M., Moroni M.;
RL Submitted (MAR-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL: U95466; AAB52813.1; -
KW Envelope protein.
FT NON_TER 1
SQ SEQUENCE 35 AA; 3843 MW; 8A05BA9090C7B2D1 CRC64;

Query Match 100.0%; Score 21; DB 14; Length 35;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DTSG 4
|||
Db 12 DTSG 15

RESULT 10
O9TSA3 PRELIMINARY: PRT: 36 AA.
AC O9TSA3.
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE F-ACTIN BINDING PROTEIN-ANNEXIN VI HOMOLOG (FRAGMENTS).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCB1_TaxID=9913;
RN [1]
RP SEQUENCE.
RX MEDLINE=93046733; PubMed=1423665;
RA Hosoya H., Kobayashi R., Tsukita S., Matsumura F.;
RT "Ca(2+)-regulated actin and phospholipid binding protein (68 kD-protein) from bovine liver: identification as a homologue for annexin VI and intracellular localization.";
RL Cell Motil. Cytoskeleton 22:200-210(1992).
DR HSSP: P79134; IAVC.
FT NON_TER 1
FT NON_CONS 20 21
FT NON_TER 36 36
SQ SEQUENCE 36 AA; 4074 MW; 051B40B5CD30FCE8 CRC64;

Query Match 100.0%; Score 21; DB 6; Length 36;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DTSG 4
||||
DB 13 DTSG 16

RESULT 11

O9TS64 PRELIMINARY; PRT; 41 AA.
AC O9TS64. (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE 33 KDA ANNEXIN V HOMOLOG (FRAGMENT).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE.
RX MEDLINE=96091481; PubMed=7585627;
RA Wang L., Rahman M.M., Iida H., Inai T., Kawabata S., Iwanaga S., Shibata Y.;
RT "Annexin V is localized in association with z-line of rat cardiac myocytes";
RL Cardiovasc. Res. 30:363-371(1995).
DR HSP; P08758; IHVF.
FT NON_TER 1
FT NON_CONS 21 22
FT NON_TER 41 41
SQ SEQUENCE 41 AA; 4575 MW; 34A58B9AC03B040 CRC64;

Query Match 100.0%; Score 21; DB 6; Length 41;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 DTSG 4
||||
DB 18 DTSG 21

RESULT 12
O9NIC2 PRELIMINARY; PRT; 45 AA.
AC O9NIC2. (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Chapman K.L., Freeman B., Freemont A.J., Hilarby M.C., Grant M.E., Boot-Handford R.P., Wallis G.A.;
RT "Differential expression of matrix GLA protein, alpha enolase, and annexin V within the epiphyseal growth plate and in human osteoarthritic tissue";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF210362; AAF25883.1;
DR InterPro; IPR001464;
DR SMART; SM00335; ANX; 1.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 45 AA; 5106 MW; 741497AFD3C8DBA5 CRC64;

Query Match 100.0%; Score 21; DB 6; Length 45;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 DTSG 4
||||
DB 27 DTSG 30

RESULT 13

O83190 PRELIMINARY; PRT; 48 AA.
AC O83190. (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE GLYCOPROTEIN B (FRAGMENT).
GN GB.
OS Mouse cytomegalovirus 1.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Betaherpesvirinae; Muromegalovirus.
OX NCBI_TaxID=10366;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=G3E;
RX MEDLINE=92292266; PubMed=1318410;
RA Rapp M., Messerle M., Buhler B., Tanneheimer M., Keil G.M., Koszinowski U.H.;
RT "Identification of the murine cytomegalovirus glycoprotein B gene and its expression by recombinant vaccinia virus";
RL J. Virol. 66:4399-4406(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=G3E;
RX MEDLINE=96145134; PubMed=8558128;
RA Xu J., Lyons P.A., Carter M.D., Booth T.W., Davis-Poynter N.J., Shellam G.R., Scalzo A.A.;
RT "Assessment of antigenicity and genetic variation of glycoprotein B of murine cytomegalovirus";
RL J. Gen. Virol. 77:49-59(1996).
DR EMBL; L39228; AAB01382.1;
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 48 AA; 5498 MW; 337DA81426D21764 CRC64;

Query Match 100.0%; Score 21; DB 14; Length 48;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DTSG 4
||||
DB 32 DTSG 35

RESULT 14
O83195 PRELIMINARY; PRT; 48 AA.
AC O83195. (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE GLYCOPROTEIN B (FRAGMENT).
GN GB.
OS Mouse cytomegalovirus 1.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Betaherpesvirinae; Muromegalovirus.
OX NCBI_TaxID=10366;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=G4;
RX MEDLINE=92292266; PubMed=1318410;
RA Rapp M., Messerle M., Buhler B., Tanneheimer M., Keil G.M., Koszinowski U.H.;
RT "Identification of the murine cytomegalovirus glycoprotein B gene and its expression by recombinant vaccinia virus";

RL J. Virol. 66:4399-4406(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-G4;
 RX MEDLINE-96145134; PubMed-8558128;
 RA Xu J., Lyons P.A., Carter M.D., Booth T.W., Davis-Poynter N.J.,
 RA Shellam G.R., Scalzo A.A.;
 RT "Assessment of antigenicity and genetic variation of glycoprotein B of
 RT murine cytomegalovirus.";
 RL J. Gen. Virol. 77:49-59(1996).
 DR EMBL: L39221; AAB01386.1; -.
 FT NON_TER 1 1
 SQ SEQUENCE 48 AA; 5498 MW; 337DA81426D21764 CRC64;

Query Match 100.0%; Score 21; DB 14; Length 48;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DTSG 4
 DB 32 DTSG 35

RESULT 15
 ID 083200 PRELIMINARY; PRT: 48 AA.
 AC 083200;
 DT 01-NOV-1996 (TRENBLREL. 01, Created)
 DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)
 DT 01-NOV-1998 (TRENBLREL. 08, Last annotation update)
 DE GLYCOPROTEIN B (FRAGMENT).
 GN GB.
 OS Mouse cytomegalovirus 1.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Betaherpesvirinae; Murinegaleovirus.
 OX NCBI_TaxID=10366;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-G5;
 RX MEDLINE-92292266; PubMed-1318410;
 RA Rapp M., Messerle M., Buhler B., Tannheimer M., Kell G.M.,
 RA Koszinowski U.H.;
 RT "Identification of the murine cytomegalovirus glycoprotein B gene and
 RT its expression by recombinant vaccinia virus.";
 RL J. Virol. 66:4399-4406(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-G5;
 RX MEDLINE-96145134; PubMed-8558128;
 RA Xu J., Lyons P.A., Carter M.D., Booth T.W., Davis-Poynter N.J.,
 RA Shellam G.R., Scalzo A.A.;
 RT "Assessment of antigenicity and genetic variation of glycoprotein B of
 RT murine cytomegalovirus.";
 RL J. Gen. Virol. 77:49-59(1996).
 DR EMBL: L39220; AAB01390.1; -.
 FT NON_TER 1 1
 SQ SEQUENCE 48 AA; 5498 MW; 337DA81426D21764 CRC64;

Query Match 100.0%; Score 21; DB 14; Length 48;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DTSG 4
 DB 32 DTSG 35

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:51:07 ; Search time 72.75 seconds

(without alignments)
1.883 Million cell updates/sec

Title: US-09-603-713-8

Perfect score: 21
Sequence: 1 DTSG 4

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	100.0	58	1	CIT2_SALTY
2	21	100.0	64	1	PSAE_GUTH
3	21	100.0	87	1	SYR_SALTY
4	21	100.0	87	1	Y14B_BPT4
5	21	100.0	101	1	RNST_SACER
6	21	100.0	122	1	VE10_LAMBD
7	21	100.0	135	1	RS6E_HALMA
8	21	100.0	137	1	PSP2_PIG
9	21	100.0	137	1	RS12_STRAU
10	21	100.0	152	1	FGF1_PIG
11	21	100.0	153	1	SODC_DEBHA
12	21	100.0	161	1	PRSA_ECOLI
13	21	100.0	163	1	GSPK_AERHY
14	21	100.0	166	1	CDN7_HUMAN
15	21	100.0	169	1	MOAB_ECOLI
16	21	100.0	171	1	IPP1_RAT
17	21	100.0	173	1	NU6M_CROLA
18	21	100.0	173	1	NU6M_ONCWA
19	21	100.0	173	1	NU6H_SALSA
20	21	100.0	181	1	FLA4_HALSA
21	21	100.0	189	1	BCP_PEA
22	21	100.0	190	1	UCRI_PARDE
23	21	100.0	195	1	YAGZ_ECOLI
24	21	100.0	196	1	FLA4_HALHA
25	21	100.0	196	1	FLA4_HALNI
26	21	100.0	197	1	MCS_MOUSE
27	21	100.0	197	1	SDC4_CHICK
28	21	100.0	202	1	LIF_BOVIN
29	21	100.0	202	1	LIF_HUMAN
30	21	100.0	202	1	LIF_MUSVI
31	21	100.0	204	1	FCAL_TRYRA
32	21	100.0	204	1	PRTB_SCYLI
33	21	100.0	205	1	DMSB_HAEIN

34	21	100.0	205	1	NU6M_ARATH	Q01825 arabidopsis
35	21	100.0	214	1	VT11_SCHPO	P78768 schizosacch
36	21	100.0	216	1	GYRB_ACIS6	O44277 acinetobact
37	21	100.0	216	1	GYRB_ACIS9	O44276 acinetobact
38	21	100.0	216	1	R11C_LORTA	O40193 locus japon
39	21	100.0	217	1	RB1A_ARATH	O96283 arabidopsis
40	21	100.0	219	1	YLP4_ZYMOO	O66114 zymomonas m
41	21	100.0	222	1	CICA_CAUCR	O87207 caulobacter
42	21	100.0	225	1	US07_HCMVA	P09731 human cytom
43	21	100.0	227	1	RHO2_HUMAN	P52198 homo sapien
44	21	100.0	229	1	RUBB_BACSU	P35159 bacillus su
45	21	100.0	229	1	YOKR_YERRU	O87571 yersinia ru

ALIGNMENTS

```

RESULT 1
CIT2_SALTY
ID CIT2_SALTY STANDARD; PRT; 58 AA.
AC P37463;
DT 01-OCT-1994 (rel. 30, Last sequence update)
DT 01-OCT-1994 (rel. 30, Last sequence update)
DT 01-NOV-1997 (rel. 35, Last annotation update)
DE CITRATE UTILIZATION PROTEIN B (FRAGMENT).
GN CITB.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=LT2;
RX MEDLINE=92041761; PubMed=1718953;
RA Shimamoto T., Izawa H., Daimon H., Ishiguro N., Shinagawa M.,
RA Sakano T., Tsuda M., Tsuchiya T.;
RT Cloning and nucleotide sequence of the gene (citA) encoding a
RT citrate carrier from Salmonella typhimurium.
RT J. Biochem. 110:22-28(1991).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; D90203; BA014227.1; -
CC DR PIR; P00098; P00098.
CC DR Styrge; SG10059; citB.
CC FT citrate utilization.
CC KW NON_TER
CC SQ SEQUENCE 58 AA; 6271 MW; 1639B1EB009A54A2 CRC64;
Query Match 100.0%; Score 21; DB 1; Length 58;
Best Local Similarity 100.0%; Pred No. 39;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 DTSG 4
Db 3 DTSG 6
RESULT 2
PSAE_GUTH
ID PSAE_GUTH STANDARD; PRT; 64 AA.
AC 078515;
DT 15-DEC-1998 (rel. 37, Last sequence update)
DT 15-DEC-1998 (rel. 37, Last sequence update)
DT 15-JUL-1999 (rel. 38, Last annotation update)
DE PHOTOSYSTEM I REACTION CENTRE SUBUNIT IV (Psl-E).

```

GN PSAE.
 OS Guillardia theta (Cryptomonas phi).
 CC Chloroplast.
 CC Eukaryota: Cryptophyta: Cryptomonadaceae; Guillardia.
 OX NCBI_TaxID=55529;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99128221; PubMed=9929392;
 RA Douglas S.E., Penny S.L.;
 RT "The plastid genome of the cryptophyte alga, Guillardia theta:
 RT complete sequence and conserved syntenic groups confirm its common
 RT ancestry with red algae.";
 RL J. Mol. Evol. 48:236-244(1999).
 CC -1- FUNCTION: MAY FORM COMPLEXES WITH FERREDOXIN AND FERREDOXIN-
 CC OXIDOREDUCTASE IN PHOTOSYSTEM I (PS I) REACTION CENTRE.
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST THYLAKOID MEMBRANE.
 CC -1- SIMILARITY: BELONGS TO THE PSAE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF041466; AAC35737.1;
 KM Photosynthesis; Photosystem I; Chloroplast; Thylakoid membrane.
 SQ SEQUENCE 64 AA: 7372 MW: 87C8C3F3D33054C2 CRC64;

Query Match 100.0%; Score 21; DB 1; Length 64;
 Best Local Similarity: 100.0%; Pred. No. 43;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 DTSG 4
 DB 28 DTSG 31

RESULT 3
 SYR_SALTY STANDARD: PRT; 87 AA.
 ID SYR_SALTY
 AC P74871;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE 01-OCT-2000 (Rel. 40, Last annotation update)
 DE ARGINYL-TRNA SYNTHETASE (EC 6.1.1.19) (ARGININE--TRNA LIGASE) (ARGRS)
 GN ARGS.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN-SL1344;
 RX MEDLINE=97084577; PubMed=8930920;
 RA Valdivia R.H., Falkow S.;
 RT "Bacterial genetics by flow cytometry: rapid isolation of Salmonella
 RT typhimurium acid-inducible promoters by differential fluorescence
 RT induction.";
 RL Mol. Microbiol. 22:367-378(1996)
 CC -1- CATALYTIC ACTIVITY: ATP + L-ARGININE + TRNA(ARG) = AMP +
 CC PYROPHOSPHATE + L-ARGINYL-TRNA(ARG).
 CC -1- SUBUNIT: MONOMER.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M22677; AAA32530.1;
 KM Hypothetical protein.
 SQ SEQUENCE 87 AA: 10065 MW: 05D0BDDF027FC295 CRC64;

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U62714; AAC44610.1;
 DR StvGene; SG10599; args.
 DR InterPro: IPR001412;
 DR PROSITE: PS00178; AA-TRNA LIGASE I; PARTIAL.
 KM Aminoacyl-tRNA synthetase: Protein biosynthesis; Ligase; ATP-binding.
 FT NON_TER
 SQ SEQUENCE 87 AA: 9724 MW: 29BC0C427F1DB90B CRC64;

Query Match 100.0%; Score 21; DB 1; Length 87;
 Best Local Similarity: 100.0%; Pred. No. 60;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 DTSG 4
 DB 77 DTSG 80

RESULT 4
 Y14B_BP74 STANDARD: PRT; 87 AA.
 ID Y14B_BP74
 AC P32280;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DE 01-FEB-1996 (Rel. 33, Last annotation update)
 DE HYPOTHETICAL 10.1 KDA PROTEIN IN NRDA-TD INTERGENIC REGION.
 GN Y14B OR NRDA.2 OR Y.
 OS Bacteriophage T4.
 OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Myoviridae;
 OC T4-like phages.
 OX NCBI_TaxID=10665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88112857; PubMed=3322944;
 RA Chu F.K., Maley G.F., Wang A.M., Maley F.;
 RT "Localization of the T4 phage ribonucleotide reductase B1 subunit
 RT gene and the nucleotide sequence of its upstream and 5' coding
 RT regions.";
 RL Gene 57:143-148(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89034091; PubMed=2846540;
 RA Tseng M.-J., Hillfinger J.M., Walsh A., Greenberg G.R.;
 RT "Total sequence, flanking regions, and transcripts of bacteriophage
 RT T4 rnda gene, coding for alpha chain of ribonucleoside diphosphate
 RT reductase.";
 RL J. Biol. Chem. 263:16242-16251(1988).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M22677; AAA32530.1;
 DR EMBL: J03968; AAA32526.1;
 KM Hypothetical protein.
 SQ SEQUENCE 87 AA: 10065 MW: 05D0BDDF027FC295 CRC64;

Query Match 100.0%; Score 21; DB 1; Length 87;
 Best Local Similarity: 100.0%; Pred. No. 60;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 DTSG 4
 DB 7 DTSG 10

```

RESULT 5
RNST_SACER STANDARD: PRT: 101 AA.
AC P00650:
DT 21-JUL-1986 (rel. 01, Created)
DT 21-JUL-1986 (rel. 01, Last sequence update)
DE 01-AUG-1992 (rel. 23, Last annotation update)
DE GUANYL-SPECIFIC RIBONUCLEASE SF (EC 3.1.27.3).
OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria, Firmicutes, Actinobacteria, Actinobacteriales,
OC Actinomycetales, Pseudonocardiaceae, Pseudonocardiaceae,
OC Saccharopolyspora.
OC NCBI_TaxID=1836;
RN [1]
RP SEQUENCE.
RX MEDLINE=76188019; PubMed=1269746;
RA Yoshida N., Sasaki A., Rashid M.A., Otsuka H.;
RT "The amino acid sequence of ribonuclease St. ";
RL FEBS Lett. 64:122-125(1976).
RN [2]
RP REVISION.
RA Yoshida N., Sasaki A., Rashid M.A., Otsuka H.;
RL Submitted (JUN-1977) to the PIR data bank.
CC -1- CATALYTIC ACTIVITY: TWO-STAGE ENDONUCLEOLYTIC CLEAVAGE TO 3'-
CC PHOSPHOMONONUCLEOTIDES AND 3'-PHOSPHODINUCLEOTIDES ENDING IN
CC G-P WITH 2',3'-CYCLIC PHOSPHATE INTERMEDIATES.
DR PIR: A00797; NRSWTE.
DR HSSP: P05798; 2SAR.
DR InterPro: IPR000026; -.
DR Pfam: PF00545; ribonuclease: 1.
DR Hydrolase: Nuclease; Endonuclease.
FT DISULFID 4 54
FT ACT_SITE 61 61 BY SIMILARITY.
FT ACT_SITE 76 76 BY SIMILARITY.
FT ACT_SITE 91 91 BY SIMILARITY.
SQ SEQUENCE 101 AA: 11354 MW; E739ADB75A61A9FB CRC64;

Query Match 100.0%; Score 21; DB 1; Length 101;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTSG 4
DB 6 DTSG 9

RESULT 6
VE10_LAMB STANDARD: PRT: 122 AA.
ID VE10_LAMB
AC P03757:
DT 21-JUL-1986 (rel. 01, Created)
DT 21-JUL-1986 (rel. 01, Last sequence update)
DT 01-NOV-1988 (rel. 09, Last annotation update)
DE EA10 GENE PROTEIN.
GN EA10 OR SSB.
OS Bacteriophage Lambda.
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae;
OC Lambda phage group.
OX NCBI_TaxID=10710;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83189071; PubMed=6221115;
RA Sanger F., Coulson A.R., Hong G.F., Hill D.F., Petersen G.B.;
RT "Nucleotide sequence of bacteriophage lambda DNA.";
RL J. Mol. Biol. 162:729-773(1982).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=82059489; PubMed=6458018;
RA Inelchen K., Shepherd J.C.W., Bickle T.A.;
RT "The DNA sequence of the phage lambda genome between PL and the gene

```

```

RT bet.";
RL Nucleic Acids Res. 9:4639-4653(1981).
CC -1- FUNCTION: THE EA10 GENE PROTEIN IS BELIEVED TO BE INVOLVED IN THE
CC PRODUCTION OF THE TRO PHENOTYPE. THIS PHENOTYPE IS EXPRESSED
CC WHEN PHAGES THAT POSSESS A MUTANT CRO GENE AND A THERMOLABILE CI
CC REPRESSOR GENE ARE UNABLE TO PROPAGATE AT RESTRICTIVE
CC TEMPERATURES. THIS INABILITY IS CORRELATED WITH THE SHUTOFF OF
CC HOST MACROMOLECULAR SYNTHESIS.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: V00638; CAA23980.1; -.
DR EMBL: J02459; AAA96574.1; -.
DR PIR: A04382; QEBPL.
SQ SEQUENCE 122 AA: 13780 MW; 9C51F65EE1118BE6 CRC64;

```

```

Query Match 100.0%; Score 21; DB 1; Length 122;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 DTSG 4
DB 107 DTSG 110

```

```

RESULT 7
RS6E_HALMA STANDARD: PRT: 135 AA.
ID RS6E_HALMA
AC P21509;
DT 01-MAY-1991 (rel. 18, Created)
DT 01-MAY-1991 (rel. 18, Last sequence update)
DT 30-MAY-2000 (rel. 39, Last annotation update)
DE 30S RIBOSOMAL PROTEIN S6E (HS13).
GN RS6E.
OS Halorcula marismortui (Halobacterium marismortui).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae; Halorcula.
OX NCBI_TaxID=2238;
RN [1]
RP SEQUENCE.
RX MEDLINE=89248680; PubMed=2655851;
RA Kimura M., Arndt E., Hatakeyama T., Kimura J.;
RT "Ribosomal proteins in halobacteria.";
RL Can. J. Microbiol. 35:195-199(1989).
CC -1- SIMILARITY: BELONGS TO THE S6E FAMILY OF RIBOSOMAL PROTEINS.
DR PIR: S11593; R3HS13.
DR InterPro: IPR001377; -.
DR Pfam: PF01092; Ribosomal_S6e; 1.
DR PROSITE: PS00578; RIBOSOMAL_S6e; 1.
KW Ribosomal protein.
SQ SEQUENCE 135 AA: 14185 MW; 8C6701F6B9271FB CRC64;

```

```

Query Match 100.0%; Score 21; DB 1; Length 135;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 DTSG 4
DB 69 DTSG 72

```

```

RESULT 8
PSP2_PIG STANDARD: PRT: 137 AA.
ID PSP2_PIG
AC P35496;
DT 01-JUN-1994 (rel. 29, Created)

```

DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE MAJOR SEMINAL PLASMA GLYCOPROTEIN PSP-II PRECURSOR.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_Taxid=9823;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Seminal vesicle;
 RA MEDLINE=94000385; PubMed=8397818;
 RA Kwok S.C.M., Yang D., Dai G., Soares M.J., Chen S., McMurtry J.P.;
 RT "Molecular cloning and sequence analysis of two porcine seminal
 RT proteins, PSP-I and PSP-II: new members of the spermadhesin family.";
 RL DNA Cell Biol. 12:605-610(1993).
 RN (2)
 RP SEQUENCE OF 22-34.
 RC TISSUE=Sperm;
 RA MEDLINE=92264734; PubMed=1586165;
 RA Rutherford K.J., Swiderex K.M., Green C.B., Chen S., Shively J.E.,
 RT "Purification and characterization of PSP-I and PSP-II, two major
 RT proteins from porcine seminal plasma.";
 RL Arch. Biochem. Biophys. 295:352-359(1992).
 RN (3)
 RP PARTIAL SEQUENCE, DISULFIDE BONDS, & CARBOHYDRATE-LINKAGE SITE N-119.
 RC TISSUE=Sperm;
 RA MEDLINE=95300970; PubMed=7781775;
 RA Calvete J.J., Mann K., Schaefer W., Raida M., Sanz L.,
 RT "Isolation and characterization of posttranslational
 RT modifications, heterodimer formation with PSP-I glycoforms and effect
 RT of dimerization on the ligand-binding capabilities of the subunits.";
 RL FEBS Lett. 365:179-182(1995).
 RN (4)
 RP STRUCTURE OF CARBOHYDRATE.
 RC MEDLINE=99435986; PubMed=10504403;
 RA Nimitz M., Grabenhorst E., Conrad H.S., Sanz L., Calvete J.J.;
 RT "Structural characterization of the oligosaccharide chains of native
 RT and crystallized boar seminal plasma spermadhesin PSP-I and PSP-II
 RT glycoforms.";
 RL Eur. J. Biochem. 265:703-718(1999).
 RN (5)
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
 RA MEDLINE=97475216; PubMed=9334740;
 RA Romero A., Romao M.J., Varela P.F., Koelln I., Dias J.M.,
 RA Carvalho A.L., Sanz L., Toepfer-Petersen E., Calvete J.J.;
 RT "The crystal structures of two spermadhesins reveal the CUB domain
 RT fold.";
 RL Nat. Struct. Biol. 4:783-788(1997).
 CC -1- SUBUNIT: MONOMER OR HETERODIMER WITH PSP-I (DEPENDING ON THE TYPE
 CC OF GLYCOSYLATION OF PSP-I).
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- TISSUE SPECIFICITY: SEMINAL PLASMA OR SPERM.
 CC -1- MASS SPECTROMETRY: MW=12410; MW_ERR=8; METHOD=MALDI.
 CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE SPERMADHESIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC EMBL: U02627; AAC48400.1;
 DR PIR: S23943; S23943.
 DR PDB: 1SP; PRELIMINARY.
 DR GlycoSuiteDB: P35496;
 DR InterPro: IPR000124;
 DR InterPro: IPR000859;
 DR Pfam: PF00431; CUB; 1.

DR PROSITE: PS00985; SPERMADHESIN_1; 1.
 DR PROSITE: PS00986; SPERMADHESIN_2; 1.
 DR PROSITE: PS01180; CUB; 1.
 KW Glycoprotein; Signal; 3D-structure.
 FT SIGNAL 1 21
 FT CHAIN 22 137
 FT DOMAIN 30 131
 FT DISULFID 30 51
 FT DISULFID 74 95
 FT CARBOHYD 119 119
 SQ SEQUENCE 137 AA; 14816 MW; 5E973910C6F97995 CRC64; N-LINKED (GLCNAC...) (COMPLEX).
 EDAC26A97DAF5D23 CRC64;
 Query Match 100.0%; Score 21; DB 1; Length 137;
 Best Local Similarity 100.0%; Pred. No. 96;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 DTSG 4
 DB 116 DTSG 119
 RESULT 10
 FGFL_PIG
 ID FGFL_PIG STANDARD; PRT; 152 AA.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC EMBL: U02869; AAC6353.1;
 DR InterPro: IPR000230;
 DR Pfam: PF00164; RIBOSOMAL_S12; 1.
 DR PRINTS: PR01034; RIBOSOMAL_S12.
 DR PROSITE: PS00055; RIBOSOMAL_S12; 1.
 KW Ribosomal protein.
 SQ SEQUENCE 137 AA; 15287 MW; 5E973910C6F97995 CRC64;
 Query Match 100.0%; Score 21; DB 1; Length 137;
 Best Local Similarity 100.0%; Pred. No. 96;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 DTSG 4
 DB 116 DTSG 119
 RESULT 10
 FGFL_PIG
 ID FGFL_PIG STANDARD; PRT; 152 AA.

```

AC P20002;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HEPARIN-BINDING GROWTH FACTOR 1 PRECURSOR (HBGF-1) (ACIDIC FIBROBLAST
DE GROWTH FACTOR) (AFGF) (ALPHA-ENDOTHELIAL CELL GROWTH FACTOR)
DE (FRAGMENT).
GN FGF1 OR FGF-1.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=92062117; PubMed=1719973;
RA Schmidt W., Sharma H.S., Schott R.J., Schaper W.;
RT "Amplification and sequencing of mRNA encoding acidic fibroblast
RT growth factor (afgf) from porcine heart."
RL Blochem. Biophys. Res. Commun. 180:853-859(1991).
RN [2]
RP SEQUENCE OF 27-41.
RX MEDLINE=89231704; PubMed=2714282;
RA Quinkler W., Maasberg W., Bernotat-Danielowski S., Luethe N.,
RA Sharma H.S., Schaper W.;
RT "Isolation of heparin-binding growth factors from bovine, porcine and
RT canine hearts."
RL Eur. J. Biochem. 181:67-73(1989).
CC -1- FUNCTION: THE HEPARIN-BINDING GROWTH FACTORS ARE ANGIOGENIC AGENTS
CC IN VIVO AND ARE POTENT MITOGENS FOR A VARIETY OF CELL TYPES IN
CC VITRO. THERE ARE DIFFERENCES IN THE TISSUE DISTRIBUTION AND
CC CONCENTRATION OF THESE 2 GROWTH FACTORS.
CC -1- SUBUNIT: MONOMER.
CC -1- MISCELLANEOUS: THIS PROTEIN BINDS HEPARIN, ALTHOUGH LESS STRONGLY
CC THAN DOES BFGF.
CC -1- SIMILARITY: BELONGS TO THE HEPARIN-BINDING GROWTH FACTORS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X60317; CAA42869.1; -
DR PIR: S03954; S03954.
DR HSSP: P05230; 2AXX.
DR InterPro: IPR002209; -
DR Pfam: PF00167; FGF_1.
DR PROSITE: PS00247; HBGF_FGF_1.
KW Growth factor; Mitogen; Vascularization; Heparin-binding.
FT PROPEP 1 15
FT CHAIN 16 >152 HEPARIN-BINDING GROWTH FACTOR 1.
FT CHAIN 22 >152 ENDOTHELIAL CELL GROWTH FACTOR ALPHA.
FT BINDING 24 28 HEPARIN (POTENTIAL).
FT BINDING 113 116 HEPARIN (POTENTIAL).
FT CONFLICT 31 31 C -> S (IN REF. 2).
FT CONFLICT 39 39 R -> Y (IN REF. 2).
FT NON_TER 152 152
SQ SEQUENCE 152 AA; 17103 MW; AEB53B0A92F9ABF4 CRC64;

```

Query Match 100.0%; Score 21; DB 1; Length 152;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 DTSG 4
 ||||
 Db .83 DTSG 86

RESULT 11

```

SODC_DEBHA
ID SODC_DEBHA STANDARD: PRT: 153 AA.
AC 042724;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE SUPEROXIDE DISMUTASE [CU-ZN] (EC 1.15.1.1).
DE SOD-1.
GN Debaromyces hansenii (Yeast) (Torulaspora hansenii).
OS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Debaromyces.
OX NCBI_TaxID=4959;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 13-28.
RC STRAIN=CIBNOR C-11;
RX MEDLINE=98265544; PubMed=9605507;
RA Hernandez-Saavedra N.Y., Egly J.-M., Ochoa J.L.;
RT "Cloning and sequencing of a cDNA encoding a copper-zinc superoxide
RT dismutase enzyme from the marine yeast Debaromyces hansenii."
RL Yeast 14:573-581(1998).
CC -1- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE
CC CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 2 PEROXIDE RADICAL + 2 H(+) -> O(2) + H(2)O(2).
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF016383; AAC50010.1; -
DR HSSP: P00445; 1SDY.
DR InterPro: IPR001424; -
DR Pfam: PF00080; sodcu; 1.
DR PRINTS: PR00068; CUZNDISMTASE.
DR PROSITE: PS00087; SOD_CU_ZN_1; 1.
DR PROSITE: PS00332; SOD_CU_ZN_2; 1.
KW Oxidoreductase; Copper; Zinc.
FT INIT_MER 0 0
FT METAL 46 46 COPPER (BY SIMILARITY).
FT METAL 48 48 COPPER (BY SIMILARITY).
FT METAL 63 63 COPPER AND ZINC (BY SIMILARITY).
FT METAL 71 71 ZINC (BY SIMILARITY).
FT METAL 80 80 ZINC (BY SIMILARITY).
FT METAL 83 83 ZINC (BY SIMILARITY).
FT METAL 120 120 COPPER (BY SIMILARITY).
FT DISULFID 57 146 BY SIMILARITY.
SQ SEQUENCE 153 AA; 15800 MW; D9349CB632EAD07 CRC64;

```

Query Match 100.0%; Score 21; DB 1; Length 153;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 DTSG 4
 ||||
 Db 90 DTSG 93

RESULT 12
 PRSA_ECOLI STANDARD: PRT: 161 AA.
 AC P42184;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE PRS FIBRILLAR MAJOR PILIN PROTEIN (PRS PILI) (FRAGMENT).
 GN PRSA.

```

OS Escherichia coli.
CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC Escherichia.
OX NCBI_TaxID=562;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=1442;
RA MEDLINE=93023852; PubMed=1357526;
RA Lindberg F., Gastera W., Noimark S., Hamers A., Baga M.,
RT "Horizontal gene transfer of the Escherichia coli pap and pps pill
RT operons as a mechanism for the development of tissue-specific
RT adhesive properties."
RT Mol. Microbiol. 6:2225-2242(1992).
CC -1- FUNCTION: FIMBRIAE (ALSO CALLED PILI), POLAR FILAMENTS RADIATING
CC FROM THE SURFACE OF THE BACTERIUM TO A LENGTH OF 0.5-1.5
CC MICROMETERS AND NUMBERING 100-300 PER CELL. ENABLE BACTERIA TO
CC COLONIZE THE EPITHELIUM OF SPECIFIC HOST ORGANS.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO THE FIMA/PAPA FAMILY OF FIMBRIA PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X62157; CAA4083.1;
CC DR InterPro: IPR000259;
CC DR Pfam: PF00419; Fimbrilal; 1.
CC Fimbrila.
CC NON TER
CC FT SEQUENCE 161 AA; 16361 MW; 1245036C9E6FBC13 CRC64;
SO

Query Match 100.0%; Score 21; DB 1; Length 161;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTSG 4
DB 139 DTSG 142

RESULT 13
GSPM_AERHY STANDARD; PRT; 163 AA.
AC P41850;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE GENERAL SECRETION PATHWAY PROTEIN M.
GN EXEM.
OS Aeromonas hydrophila.
OC Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;
OC Aeromonas.
OX NCBI_TaxID=644;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=AH65;
RA MEDLINE=94012544; PubMed=8407845;
RA Howard S.P., Critch J., Bedi A.;
RT "Isolation and analysis of eight exo genes and their involvement in
RT extracellular protein secretion and outer membrane assembly in
RT Aeromonas hydrophila."
RT J. Bacteriol. 175:6695-6703(1993).
CC -1- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE
CC EXPORT OF PROTEINS.
CC -1- SUBCELLULAR LOCATION: INNER MEMBRANE (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE EXEM/PULM/OUTM/XCF2 FAMILY.
CC -----

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X66504; CAA47134.1;
CC DR TRANSPO: Transmembrane; Inner membrane.
CC FT TRANSMEM 20 40
CC SEQUENCE 163 AA; 18571 MW; 3D3428347E081436 CRC64;
SO

Query Match 100.0%; Score 21; DB 1; Length 163;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTSG 4
DB 80 DTSG 83

RESULT 14
CDN7_HUMAN STANDARD; PRT; 166 AA.
ID CDN7_HUMAN
AC P55273; O13102;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CYCLIN-DEPENDENT KINASE 4 INHIBITOR D (P19-INK4d).
GN CDKN2D.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP TISSUE=Bone marrow;
RC MEDLINE=96121373; PubMed=8575754;
RA Okuda T., Hirai H., Valentine V.A., Shurtleff S.A., Kidd V.J.,
RA Lahti J.M., Sherr C.J., Downing J.R.;
RT "Molecular cloning, expression pattern, and chromosomal localization
RT of human CDKN2D/INK4d, an inhibitor of cyclin D-dependent kinases."
RT J. Biol. Chem. 270:623-630(1995).
RN (2)
RP SEQUENCE FROM N.A.
RC MEDLINE=96362662; PubMed=8741839;
RA Guan K.L., Jenkins C.W., Li Y., O'Keefe C.L., Noh S., Wu X.,
RA Zariwala M., Materra A.G., Xiong Y.;
RT "Isolation and characterization of p19INK4d, a p16-related inhibitor
RT specific to CDK6 and CDK4."
RT MOL. BIOL. Cell 7:57-70(1996).
RN (3)
RP SEQUENCE FROM N.A.
RC TISSUE=Thymus;
RC MEDLINE=95257949; PubMed=7739548;
RA Chan F.K.M., Zhang J., Cheng L., Shapiro D.N., Wimoto A.;
RT "Identification of human and mouse p19, a novel CDK4 and CDK6
RT inhibitor with homology to p16ink4."
RT Mol. Cell. Biol. 15:2682-2688(1995).
RN (4)
RP SEQUENCE FROM N.A.
RA Newton Bishop J.A., Harland M., Bennett D.C., Bataille V.,
RA Goldstein A.M., Tucker M.A., Ponder B.A.J., Cuzick J., Selby P.,
RT "Mutation testing in melanoma families: INK4A, CDK4 and INK4D."
RT Submitted (Apr-1998) to the EMBL/Genbank/DBJ databases.
RN (5)
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF COMPLEX WITH CDK6.
RA MEDLINE=98421670; PubMed=9751050;
RA Russo A.A., Tong L., Lee J.O., Jeffrey P.D., Pavletich N.P.;
RT "Structural basis for inhibition of the cyclin-dependent kinase Cdk6

```



```

RT by the tumour suppressor p16INK4a."
RL Nature 395:237-243(1998).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX MEDLINE-98455510; PubMed-9782052;
RA Baumgartner R., Fernandez-Catalan C., Minoto A., Huber R., Eng R.A.,
RA Holak T.A.;
RT "Structure of human cyclin-dependent kinase inhibitor p16(INK4d):
RT comparison to known ankyrin-repeat-containing structures and
RT implications for the dysfunction of tumor suppressor p16(INK4e).";
RL Structure 6:1279-1290(1998).
CC -1- FUNCTION: INTERACTS STRONGLY WITH CDK4 AND CDK6.
CC -1- SIMILARITY: BELONGS TO THE CDKN2 FAMILY OF CYCLIN-DEPENDENT KINASE
CC INHIBITORS.
CC -1- SIMILARITY: CONTAINS 4 ANK REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U49399; AAB03772.1; -.
DR EMBL: U40343; AAB18139.1; -.
DR EMBL: U20498; AAB85436.1; -.
DR EMBL: AF061327; AAC27450.1; -.
DR PDB: 1B18; 16-FEB-99.
DR PDB: 1BD8; 14-OCT-98.
DR MIM: 600927; -.
DR InterPro: IPR002110; -.
DR Pfam: PF000023; ank; 3.
DR PROSITE: PS50088; ANK_REPEAT; 1.
DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
DR Cell cycle; Anti-oncogene; Repeat; ANK repeat; 3D-structure.
KW REPEAT 41 69 ANK 1.
FT REPEAT 73 102 ANK 2.
FT REPEAT 106 135 ANK 3.
FT REPEAT 138 166 ANK 4.
FT CONFLICT 159 159 Q -> P (IN REF. 3).
SQ SEQUENCE 166 AA; 17700 MW; 2FACD11CF56340DC CRC64;

Query Match 100.0%; Score 21; DB 1; Length 166;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DTSG 4
DB 71 DTSG 74

RESULT 15
MOAB_ECOLI STANDARD: PRT; 169 AA.
ID MOAB_ECOLI
AC P30746;
DT 01-APR-1993 (rel. 25, Created)
DT 01-FEB-1994 (rel. 28, Last sequence update)
DT 30-MAY-2000 (rel. 39, Last annotation update)
DE MOLYBDENUM COFACTOR BIOSYNTHESIS PROTEIN B.
MOAB OR CHLA2.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.
RC STRAIN-K12 / MC4100;
RX MEDLINE-93368423; PubMed-8361352;
RA Rivers S.L., McNaughton E., Blasco F., Giordano G., Boxer D.H.;
RT "Molecular genetic analysis of the moa operon of Escherichia coli
RT K-12 required for molybdenum cofactor biosynthesis.";
```

```

RL MOL. Microbiol. 8:1071-1081(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE-9742617; PubMed-9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE-97061202; PubMed-8905232;
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
CC -1- FUNCTION: INVOLVED IN THE BIOSYNTHESIS OF MOLYBDOPTERIN PRECURSOR
CC 2 FROM GUANOSINE.
CC -1- PATHWAY: MOLYBDENUM COFACTOR BIOSYNTHESIS.
CC -1- INDUCTION: BY ANAEROBIOSIS, REPRESSED BY THE MOLYBDENUM COFACTOR.
CC -1- SIMILARITY: TO THE N-TERMINAL OF CINNAMON/GERPHRIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X70420; CAA49862.1; -.
DR EMBL: AE000181; AAC73869.1; -.
DR EMBL: D90716; BAA35440.1; -.
DR PIR: S31880; S31880.
DR PIR: S34999; S34999.
DR SWISS-2DPAGE: P30746; COLI.
DR Ecogene; EGI1596; moab.
DR InterPro: IPR001453; -.
DR PROSITE: PS01078; MOCE_BIOSYNTHESIS_1; 1.
KW Molybdenum cofactor biosynthesis.
FT INIT_MET 0
SQ SEQUENCE 169 AA; 18534 MW; 743F11472522FAEA CRC64;

Query Match 100.0%; Score 21; DB 1; Length 169;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DTSG 4
DB 26 DTSG 29
```

Search completed: September 6, 2001, 16:51:08
Job time: 812 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:45:50 ; Search time 134.15 Seconds
(without alignments)
2.271 Million cell updates/sec

Title: US-09-603-713-8

Perfect score: 21

Sequence: 1 DTSG 4

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR68:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	16	2	BA8406
2	21	100.0	20	2	BA8406
3	21	100.0	34	2	C31461
4	21	100.0	41	2	T01379
5	21	100.0	52	2	C38230
6	21	100.0	58	2	PC0098
7	21	100.0	80	2	C82669
8	21	100.0	85	2	T05719
9	21	100.0	90	2	E71527
10	21	100.0	93	2	T18118
11	21	100.0	96	2	G83084
12	21	100.0	100	2	PC0818
13	21	100.0	101	1	NRSMTF
14	21	100.0	104	1	JC4694
15	21	100.0	104	2	S11130
16	21	100.0	109	2	S38496
17	21	100.0	111	2	S72771
18	21	100.0	113	2	A85883
19	21	100.0	116	2	S65875
20	21	100.0	118	2	G64544
21	21	100.0	118	2	F71964
22	21	100.0	119	2	D75430
23	21	100.0	122	1	QEBPL
24	21	100.0	122	2	A53878
25	21	100.0	122	2	C85638
26	21	100.0	122	2	B85848
27	21	100.0	122	2	C70983
28	21	100.0	122	2	G82500
29	21	100.0	123	2	F83112

30	21	100.0	123	2	T20279	hypothetical prote
31	21	100.0	124	2	E81737	hypothetical prote
32	21	100.0	126	2	S49392	hypothetical prote
33	21	100.0	128	2	T35073	probable phosphori
34	21	100.0	134	2	S08991	Calcimedn, 34K -
35	21	100.0	135	1	R3HS13	ribosomal protein
36	21	100.0	135	2	H84523	probable thionin
37	21	100.0	137	2	F69436	conserved hypothet
38	21	100.0	137	2	A84333	transcription regu
39	21	100.0	138	2	S24076	envelope protein -
40	21	100.0	138	2	T04560	hypothetical prote
41	21	100.0	140	2	S64908	hypothetical prote
42	21	100.0	141	2	T43609	yopH chaperone syc
43	21	100.0	141	2	S70095	sych protein - Yer
44	21	100.0	143	2	F71024	hypothetical prote
45	21	100.0	150	2	G84260	hypothetical prote

ALIGNMENTS

RESULT 1
BA8406
annexin VI homolog - bovine (fragment)
C:Species: Bos primigenius taurus (cattle)
C:Date: 19-Nov-1993 #sequence_revision 18-Nov-1994 #text_change 03-Mar-1995
C:Accession: BA8406
R:Hosoya, H.; Kobayashi, R.; Tsukita, S.; Matsumura, F.
Cell Motil. Cytoskeleton 22, 200-210, 1992
A:Title: Ca(2+)-regulated actin and phospholipid binding protein (68 kD-protein) from
A:Accession: BA8406; MUID:93046733
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-16 <HOS>
A:Experimental source: liver
A>Note: sequence extracted from NCBI backbone (NCBIP:119426)

Query Match 100.0%; Score 21; DB 2; Length 16;

Best Local Similarity 100.0%; Pred. No. 27;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTSG 4
DB 9 DTSG 12

RESULT 2
A48406
annexin VI homolog - bovine (fragment)
C:Species: Bos primigenius taurus (cattle)
C:Date: 19-Nov-1993 #sequence_revision 18-Nov-1994 #text_change 03-Mar-1995
C:Accession: A48406
R:Hosoya, H.; Kobayashi, R.; Tsukita, S.; Matsumura, F.
Cell Motil. Cytoskeleton 22, 200-210, 1992
A:Title: Ca(2+)-regulated actin and phospholipid binding protein (68 kD-protein) from
A:Reference number: A48406; MUID:93046733
A:Accession: A48406
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-20 <HOS>
A:Experimental source: liver
A>Note: sequence extracted from NCBI backbone (NCBIP:119419)

Query Match 100.0%; Score 21; DB 2; Length 20;

Best Local Similarity 100.0%; Pred. No. 33;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTSG 4
DB 13 DTSG 16

```

RESULT 3
C31461
T-cell receptor delta chain BDN5, thymus - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 05-Oct-1999 #sequence_revision 05-Oct-1989 #text_change 30-May-1997
C:Accession: C31461
R:Liacy, M.J.; McNeil, L.K.; Roth, M.E.; Kranz, D.M.
Proc. Natl. Acad. Sci. U.S.A. 86, 1023-1026, 1989
A:Title: T-cell receptor delta-chain diversity in peripheral lymphocytes.
A:Reference number: A31461; MUID:89128840
A:Accession: C31461
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-34 <LAC>
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: T-cell receptor

Query Match          100.0%; Score 21; DB 2; Length 34;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTSG 4
    ||||
Db 14 DTSG 17

RESULT 4
T01379
E2/NS1 envelope protein - hepatitis C virus (fragment)
C:Species: hepatitis C virus
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000
C:Accession: T01379
R:Patelini, P.; Dirss, F.; Nalpas, B.; Pisi, E.; Franco, D.; Berchelot, P.; Brechot, C.
Hepatology 17, 20-29, 1993
A:Title: Persistence of hepatitis B and hepatitis C viral genomes in primary liver cancer
A:Reference number: Z14311; MUID:93138570
A:Accession: T01379
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-41 <PAT>
A:Cross-references: EMBL:S53725; NID:9264215
C:Superfamily: hepatitis C virus genome polypeptide

Query Match          100.0%; Score 21; DB 2; Length 41;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTSG 4
    ||||
Db 12 DTSG 15

RESULT 5
C38230
Inorganic pyrophosphatase (EC 3.6.1.1) 1 - beet (fragments)
C:Species: Beta vulgaris (beet)
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: C38230; E38230; F38230
R:Sarafian, V.; Kim, Y.; Poole, R.J.; Rea, P.A.
Proc. Natl. Acad. Sci. U.S.A. 89, 1775-1779, 1992
A:Title: Molecular cloning and sequence of cDNA encoding the pyrophosphate-energized vac
A:Reference number: A38230; MUID:92179265
A:Accession: C38230
A:Molecule type: protein
A:Residues: 1-16 <SARI>
A:Note: sequence extracted from NCBI backbone (NCBIP:87192)
A:Accession: E38230
A:Molecule type: protein
A:Residues: 17-28 <SAR2>

```

```

A:Note: sequence extracted from NCBI backbone (NCBIP:87194)
A:Accession: F38230
A:Molecule type: protein
A:Residues: 29-52 <SAR3>
A:Experimental source: cultivar Detroit Dark
A:Note: sequence extracted from NCBI backbone (NCBIP:87195)
C:Superfamily: inorganic pyrophosphatase, H+-translocating pyrophosphate-energized
C:Keywords: hydrolase

Query Match          100.0%; Score 21; DB 2; Length 52;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTSG 4
    ||||
Db 42 DTSG 45

RESULT 6
PQ0098
cIb protein - Salmonella typhimurium (fragment)
C:Species: Salmonella typhimurium
C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 20-Jun-2000
C:Accession: PQ0098
R:Shimamoto, T.; Izawa, H.; Daimon, H.; Ishiguro, N.; Shinagawa, M.; Sakano, Y.; Tsud
J. Biochem. 110, 22-28, 1991
A:Title: Cloning and nucleotide sequence of the gene (cIb) encoding a citrate carrie
A:Reference number: J00576; MUID:92041761
A:Accession: PQ0098
A:Molecule type: DNA
A:Residues: 1-58 <SHI>
A:Cross-references: GB:D90203; NID:9217050; PIDN:BA14227.1; PID:9217051
A:Experimental source: strain LT2
C:Genetics:
A:Gene: cIb
C:Superfamily: Escherichia coli plasmid pWR60 hypothetical 42K protein

Query Match          100.0%; Score 21; DB 2; Length 58;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTSG 4
    ||||
Db 3 DTSG 6

RESULT 7
C82669
50S ribosomal protein L31 XF1534 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 03-Nov-2000
C:Accession: C82669
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: C82669
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-80 <STM>
A:Cross-references: GB:AE003983; GB:AE003849; NID:9106567; PIDN:AAF84343.1; GSPDB:GN
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
Briones, M.R.S.; Bueno, M.R.P.; Camargo, L.E.A.; Carraro, D.M.; Carre
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to Genbank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marques, M.V.; Martins
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.

```

, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmeri, D.A.
Rodrigues, V.; Rosa, A.J.; de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A.Autours: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF1534
C:Superfamily: Escherichia coli ribosomal protein L31

Query Match 100.0%; Score 21; DB 2; Length 80;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DTSG 4
||||
DB 66 DTSG 69

RESULT 8
T05719
metallothionein II PGMPM19 [similarity] - soybean
C:Species: Glycine max (soybean)
C:Date: 09-Apr-1999 #sequence_revision 09-Apr-1999 #text_change 19-May-2000
C:Accession: T05719

R:Hisu, T.F.; Tsai, F.Y.; Hsing, Y.I.; Chow, T.Y.
submitted to the EMBL Data Library, June 1997

A:Description: Glycine max metallothionein-II protein mRNA.

A:Reference number: Z15431

A:Accession: T05719

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-85 <HSU>
A:Cross-references: EMBL:AF010186; NID:g2306978; PIDN:AAB65792.1; PID:g2306979

A:Experimental source: strain SH1-Sht; cotyledon

C:Genetics:
A:Gene: PGMPM19
C:Superfamily: metallothionein

Query Match 100.0%; Score 21; DB 2; Length 85;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DTSG 4
||||
DB 3 DTSG 6

RESULT 9
E71527
hypothetical protein CT330 - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C:Species: Chlamydia trachomatis
C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
C:Accession: E71527

R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,
Science 282, 754-759, 1998

A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trac
A:Reference number: A71570; MUID:99000809

A:Accession: E71527

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-90 <ARN>

A:Cross-references: GB:AE001306; GB:AE001273; NID:g3328748; PIDN:AAC67925.1; PID:g332874
A:Experimental source: serotype D, strain UW-3/Cx

C:Genetics:
A:Gene: CT330

Query Match 100.0%; Score 21; DB 2; Length 90;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DTSG 4
||||
DB 74 DTSG 77

RESULT 10
T18118
hypothetical protein a616r - Chlorella virus PCV-1
C:Species: Chlorella virus PCV-1
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T18118
R:Graves, M.V.; Van Etten, J.L.
submitted to the EMBL Data Library, May 1999

A:Reference number: Z18806

A:Accession: T18118

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-93 <GRA>

A:Cross-references: EMBL:U42580; NID:g4028896; PIDN:AAC97027.1

A:Experimental source: specific host Chlorella strain NC64A

C:Genetics:
A:Note: a616r

Query Match 100.0%; Score 21; DB 2; Length 93;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DTSG 4
||||
DB 28 DTSG 31

RESULT 11
G83084
Glu-tRNA(Gln) amidotransferase subunit C PA4482 [imported] - Pseudomonas aeruginosa (C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: G83084

R:Stover, C.K.; Pham, X.Q.; Ewlin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; L

.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa

A:Reference number: A82950; MUID:20437337

A:Accession: G83084

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-96 <STO>

A:Cross-references: GB:AE004862; GB:AE004091; NID:g9950716; PIDN:AAC07870.1; GSPDB:GN

A:Experimental source: strain PA01

C:Genetics:
A:Gene: galc; PA4482

Query Match 100.0%; Score 21; DB 2; Length 96;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DTSG 4
||||
DB 45 DTSG 48

RESULT 12
P00818
expressed sequence tag R80 - rape (fragment)
C:Species: Brassica napus (rape)
C:Date: 03-May-1994 #sequence_revision 07-Oct-1994 #text_change 03-May-1996

C:Accession: P00818
R:Park, Y.S.; Kwak, J.M.; Kwon, O.Y.; Kim, Y.S.; Lee, D.S.; Cho, M.J.; Lee, H.H.; Nam

Plant Physiol. 103, 359-370, 1993

A:Title: Generation of expressed sequence tags of random root cDNA clones of Brassica na
 A:Reference number: PQ0816; MUID:94302145
 A:Accession: PQ0818
 A:Molecule type: mRNA
 A:Residues: 1-100 <P>
 A:Experimental source: root, cv. Nahen
 C:Superfamily: heat shock protein 90

Query Match 100.0%; Score 21; DB 2; Length 100;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTSG 4
 ||||
 Db 28 DTSG 31

RESULT 13
 NRSMT
 ribonuclease Sa (EC 3.1.27.-) - Saccharopolyspora erythraea
 N:Alternate names: guanyloribonuclease; ribonuclease St
 C:Species: Saccharopolyspora erythraea
 C:Date: 24-Apr-1984 #sequence_revision 24-Apr-1984 #text_change 06-Dec-1996
 C:Accession: A91429; A00797
 R:Yoshida, N.; Sasaki, A.; Rashid, M.A.; Otsuka, H.
 FEBS Lett. 64, 122-125, 1976
 A:Title: The amino acid sequence of ribonuclease St.
 A:Reference number: A91429; MUID:7618019
 A:Accession: A91429
 A:Molecule type: protein
 A:Residues: 1-25, G, 26-101 <YOS>
 R:Yoshida, N.; Sasaki, A.; Rashid, M.A.; Otsuka, H.
 submitted to the Atlas, June 1977
 A:Reference number: A94589
 A:Contents: annotation; revision
 A:Note: a Gly following residue 25 has been deleted from the published sequence
 C:Superfamily: ribonuclease Sa
 C:Keywords: extracellular protein; hydrolase
 F:4-54/Disulfide bonds: #status experimental

Query Match 100.0%; Score 21; DB 1; Length 101;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTSG 4
 ||||
 Db 6 DTSG 9

RESULT 14
 JC4694
 proteic killer active protein hlg B - plasmid Rts1
 C:Species: plasmid Rts1
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
 C:Accession: JC4694
 R:Tian, Q.B.; Ohnishi, M.; Tabuchi, A.; Terawaki, Y.
 Biochem. Biophys. Res. Commun. 220, 280-284, 1996
 A:Title: A new plasmid-encoded proteic killer gene system: Cloning, sequencing, and alal
 A:Reference number: JC4693; MUID:96184644
 A:Accession: JC4694
 A:Molecule type: DNA
 A:Residues: 1-104 <TIA>
 A:Cross-references: GB:U43847; NID:g1262203; PID:AC43983.1; PID:g1262205
 C:Genetics:
 A:Gene: hlgB
 A:Genome: plasmid
 C:Superfamily: hypothetical protein b1477

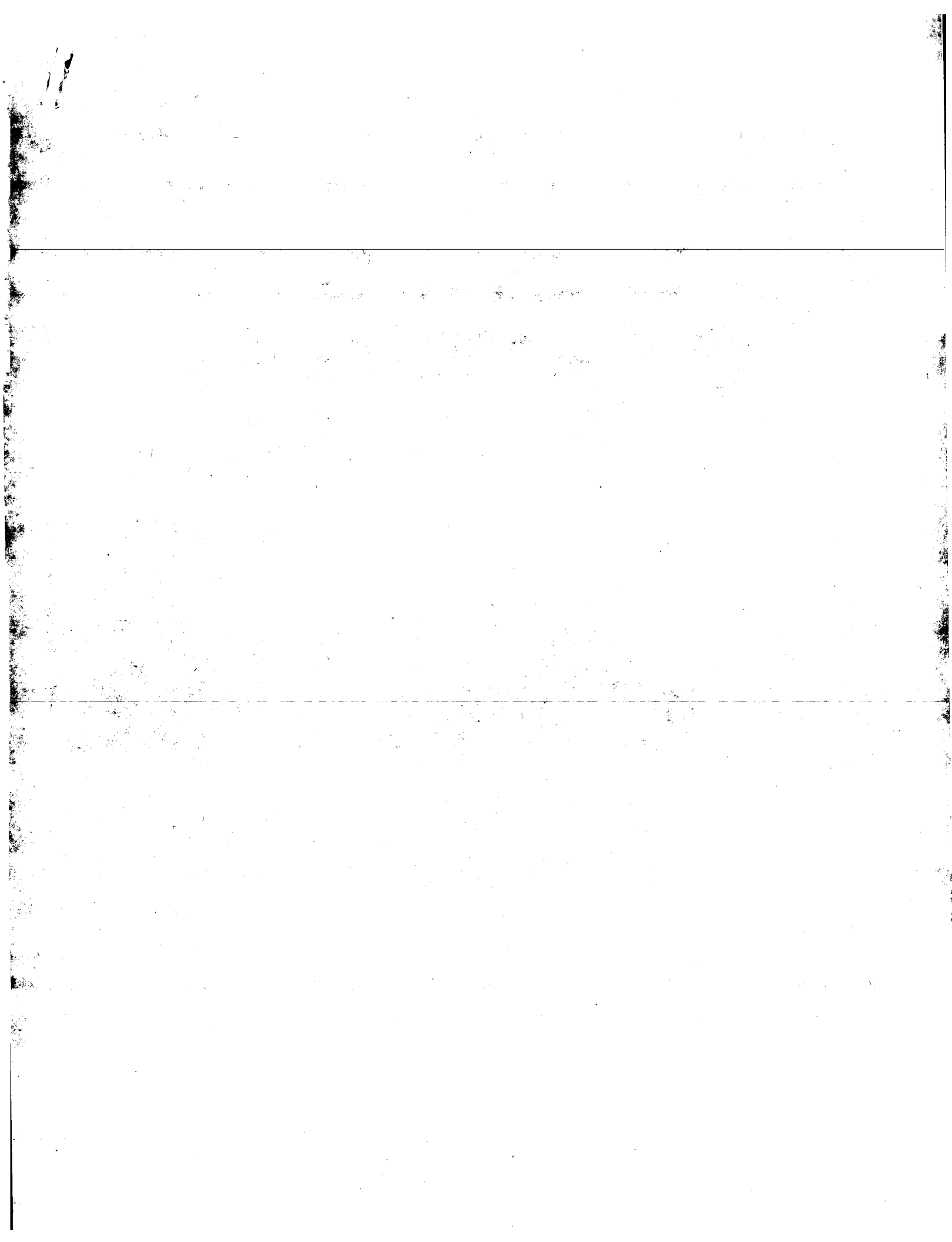
*Query Match 100.0%; Score 21; DB 1; Length 104;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DTSG 4
 ||||
 Db 84 DTSG 87

RESULT 15
 S1130
 glycoprotein 118 BC - Trypanosoma brucei
 C:Species: Trypanosoma brucei
 C:Date: 18-Feb-1994 #sequence_revision 01-Dec-1995 #text_change 07-Feb-1997
 C:Accession: S1130
 R:Lin, A.Y.C.; van der Ploeg, L.H.T.; Rijsewijk, F.A.M.; Borst, P.
 J. Mol. Biol. 167, 57-75, 1983
 A:Title: The transposition unit of variant surface glycoprotein gene 118 of Trypanoso
 A:Reference number: S07254; MUID:83241749
 A:Accession: S1130
 A:Status: preliminary; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-104 <LIU>
 A:Cross-references: EMBL:V01550
 C:Keywords: glycoprotein

Query Match 100.0%; Score 21; DB 2; Length 104;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DTSG 4
 ||||
 Db 78 DTSG 81

Search completed: September 6, 2001, 16:45:50
 Job time: 494 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:39:36 ; Search time 113.12 Seconds
(without alignments)
0.728 Million cell updates/sec

Title: US-09-603-713-8

Perfect score: 21

Sequence: 1 DTSC 4

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 1008
Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCrUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfilest1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	100.0	15	1	US-08-218-025A-159
2	21	100.0	21	1	US-08-445-135-7
3	21	100.0	25	1	US-08-345-527-6
4	21	100.0	25	3	US-08-887-769-6
5	21	100.0	28	2	US-08-968-542C-32
6	21	100.0	30	2	US-08-143-311B-9
7	21	100.0	37	2	US-09-066-074-14
8	21	100.0	37	2	US-08-555-912A-14
9	21	100.0	37	3	US-08-348-518C-24
10	21	100.0	37	3	US-08-476-509B-24
11	21	100.0	54	4	US-09-188-930-332
12	21	100.0	66	1	US-08-475-213-3
13	21	100.0	109	4	US-09-276-071-2
14	21	100.0	124	1	US-08-478-039-78
15	21	100.0	124	1	US-08-476-349A-78
16	21	100.0	158	2	US-08-378-617A-32
17	21	100.0	158	2	US-08-403-852D-29
18	21	100.0	158	3	US-08-510-646B-30
19	21	100.0	158	4	US-09-231-818-29
20	21	100.0	166	3	US-08-384-106A-5
21	21	100.0	166	3	US-08-384-106A-20
22	21	100.0	166	3	US-08-384-106A-21
23	21	100.0	166	4	US-09-240-906-2
24	21	100.0	166	5	PCR-US96-01643-9
25	21	100.0	166	4	PCR-US96-05252-2
26	21	100.0	172	4	US-08-149-101A-21
27	21	100.0	172	5	PCR-US94-12873-21

28	21	100.0	174	4	US-08-149-101A-20	Sequence 20, Appl
29	21	100.0	174	5	PCR-US94-12873-20	Sequence 20, Appl
30	21	100.0	179	1	US-08-076-087-1	Sequence 1, Appl
31	21	100.0	179	1	US-08-411-224-1	Sequence 1, Appl
32	21	100.0	179	3	US-08-911-321-3	Sequence 3, Appl
33	21	100.0	179	5	PCR-US91-07759-1	Sequence 1, Appl
34	21	100.0	181	4	US-08-505-187-2	Sequence 2, Appl
35	21	100.0	181	4	US-08-505-187-3	Sequence 3, Appl
36	21	100.0	188	2	US-08-933-750C-8	Sequence 8, Appl
37	21	100.0	188	4	US-09-234-613-8	Sequence 8, Appl
38	21	100.0	192	4	US-09-276-071-4	Sequence 4, Appl
39	21	100.0	202	1	US-08-792-019B-14	Sequence 14, Appl
40	21	100.0	202	3	US-09-106-182-5	Sequence 5, Appl
41	21	100.0	202	3	US-08-988-819-14	Sequence 14, Appl
42	21	100.0	202	4	US-09-016-534-14	Sequence 14, Appl
43	21	100.0	202	4	US-08-097-869-1	Sequence 1, Appl
44	21	100.0	205	3	US-08-549-515-6	Sequence 6, Appl
45	21	100.0	205	3	US-08-549-515-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-08-218-025A-159
Sequence 159, Application US/08218025A
Patent No. 5556744
GENERAL INFORMATION:
APPLICANT: Welner, David B.
APPLICANT: Ugen, Kenneth E.
APPLICANT: Williams, William V.
TITLE OF INVENTION: Methods and Compositions for Diagnosing
TITLE OF INVENTION: and Treating Certain HIV Infected Patients
NUMBER OF SEQUENCES: 197
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: P.O. Box 457, 321 No. 55567441sttown Road
CITY: Spring House
STATE: Pennsylvania
COUNTRY: U.S.A.
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/218,025A
FILING DATE: 24-MAR-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/891,451
FILING DATE: 29-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: WST33A
TELEPHONE: (215) 540-9206
TELEFAX: (215) 540-5818
INFORMATION FOR SEQ ID NO: 159:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-218-025A-159

Query Match 100.0% Score 21; DB 1; Length 15;
Best Local Similarity 100.0% Pred. NO. 19;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DTSG 4
DB 1 DTSG 4

RESULT 2

US-08-445-135-7
Sequence 7, Application US/08445135
Patent No. 5558789
GENERAL INFORMATION:
INVENTOR: Scott W. Rogers, Lorise C. Gahring, Roy E.
APPLICANT: Scott W. Rogers, Lorise C. Gahring, Roy E.
TITLE OF INVENTION: Promotion of Epithelial Cell Adhesion
and Hemidesmosome Assembly by a Laminin-Like Molecule
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESSES:
ADDRESS: 520 Newport Center Drive, Sixteenth Floor
CITY: Newport Beach
STATE: CA
COUNTRY: USA
ATTORNEY/AGENT INFORMATION:
NAME: Alan J. Howarth
REGISTRATION/DOCKET NUMBER: 36,553
TELECOMMUNICATION INFORMATION:
TELEPHONE: (801) 566-0750
TELEFAX: (801) 566-0750
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids
TOPOLOGY: linear
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/151,134
FILING DATE: 12 NOV -1993
ATTORNEY/AGENT INFORMATION:
NAME: Israelish, Nand A.
REGISTRATION/DOCKET NUMBER: 29,655
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 123-0175
TELEFAX: (619) 123-0175
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TOPOLOGY: linear
MOLECULE TYPE: Peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: Internal
US-08-445-135-7

Query Match 100.0%; Score 21; DB 1; Length 21;
Best Local Similarity 100.0%; Pred. No. 27;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DTSG 4
DB 17 DTSG 20

RESULT 3

US-08-345-527-6
Sequence 6, Application US/08345527
Patent No. 5731410
GENERAL INFORMATION:
INVENTOR: Scott W. Rogers, Lorise C. Gahring, Roy E.
APPLICANT: Scott W. Rogers, Lorise C. Gahring, Roy E.
TITLE OF INVENTION: Autoantibodies to Neurotransmitter
Receptors
NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:
ADDRESSER: Thorpe No. 5731410th & Western
STREET: 9035 South 700 East, Suite 200
CITY: Sandy
STATE: Utah
COUNTRY: USA
ATTORNEY/AGENT INFORMATION:
NAME: Alan J. Howarth
REGISTRATION/DOCKET NUMBER: 36,553
TELECOMMUNICATION INFORMATION:
TELEPHONE: (801) 566-0750
TELEFAX: (801) 566-0750
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids
TOPOLOGY: linear
US-08-345-527-6

Query Match 100.0%; Score 21; DB 1; Length 25;
Best Local Similarity 100.0%; Pred. No. 32;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DTSG 4
DB 18 DTSG 21

RESULT 4

US-08-887-769-6
Sequence 6, Application US/08887769
Patent No. 5731410
GENERAL INFORMATION:
INVENTOR: Scott W. Rogers, Lorise C. Gahring, Roy E.
APPLICANT: Scott W. Rogers, Lorise C. Gahring, Roy E.
TITLE OF INVENTION: Autoantibodies to Neurotransmitter
Receptors
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESSES:
ADDRESS: 520 Newport Center Drive, Sixteenth Floor
CITY: Newport Beach
STATE: CA
COUNTRY: USA
ATTORNEY/AGENT INFORMATION:
NAME: Alan J. Howarth
REGISTRATION/DOCKET NUMBER: 36,553
TELECOMMUNICATION INFORMATION:
TELEPHONE: (801) 566-0750
TELEFAX: (801) 566-0750
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids
TOPOLOGY: linear
US-08-887-769-6

NAME: Alan J. Howarth
REGISTRATION NUMBER: 36,553
REFERENCE/DOCKET NUMBER: T2411
TELECOMMUNICATION INFORMATION:
TELEPHONE: (801)566-6633
TELEFAX: (801)566-0750
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-887-769-6

Query Match 100.0%; Score 21; DB 3; Length 25;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 DTSG 4
1111
Db 18 DTSG 21

RESULT 5
US-08-968-542C-32
Sequence 32, Application US/08968542C
Patent No. 5981728
GENERAL INFORMATION:
APPLICANT: Myers, et al.
TITLE OF INVENTION: dult Codes For A No. 5981728el Starch
TITLE OF INVENTION: Synthase
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Mcgregor & Adler, LLP
STREET: 8011 Candle Lane
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh
SOFTWARE: Microsoft Word 6.0.1 for Macintosh
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/968,542C
FILING DATE: No. 5981728ember 12, 1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Benjamin Aaron Adler, Ph.D., J.D.
REGISTRATION NUMBER: 35,423
REFERENCE/DOCKET NUMBER: D6036
TELECOMMUNICATION INFORMATION:
TELEPHONE: (713) 777-2321
TELEFAX: (713) 777-6908
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 28 amino acid residues
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: amino acid
DESCRIPTION: no
HYPOTHETICAL: no
ANTI-SENSE: no
FRAGMENT TYPE: internal
US-08-968-542C-32
Query Match 100.0%; Score 21; DB 2; Length 28;
Best Local Similarity 100.0%; Pred. No. 37;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 DTSG 4
1111
Db 16 DTSG 19

RESULT 6
US-08-143-311B-9
Sequence 9, Application US/08143311B
Patent No. 5863540
GENERAL INFORMATION:
APPLICANT: HAYNES, BARTON F.
APPLICANT: HALE, LAURA P.
APPLICANT: PATTON, KAREN L.
APPLICANT: TELEN, MARILYN J.
APPLICANT: LIAO, HUA-XIN
TITLE OF INVENTION: AN ADHESION MOLECULE
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHAYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/143,311B
FILING DATE: 29-OCT-1993
CLASSIFICATION: 436
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,339
FILING DATE: 30-OCT-1992
CLASSIFICATION: 436
APPLICATION NUMBER: 07/669,730
FILING DATE: 15-MAR-1991
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: WILSON, MARY J.
REGISTRATION NUMBER: 32,955
REFERENCE/DOCKET NUMBER: 1579-50
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-816-4000
TELEFAX: 703-816-4100
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-143-311B-9
Query Match 100.0%; Score 21; DB 2; Length 30;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 DTSG 4
1111
Db 20 DTSG 23
RESULT 7
US-09-066-074-14
Sequence 14, Application US/09066074
Patent No. 5952467
GENERAL INFORMATION:

APPLICANT: Hunter et al., Tony
TITLE OF INVENTION: NIMA INTERACTING PROTEINS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/066,074
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/555,912
FILING DATE: 13-NOV-1995
ATTORNEY/AGENT INFORMATION:
NAME: Halle, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07251/011001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 37 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
IMMEDIATE SOURCE:
CLONE: FE65/Rat
US-09-066-074-14

Query Match 100.0%; Score 21; DB 2; Length 37;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 DTSG 4
1111
Db 12 DTSG 15

RESULT 8
US-08-555-912A-14
Sequence 14, Application US/0855912A
Patent No. 5972697
GENERAL INFORMATION:
APPLICANT: Hunter et al., Tony
TITLE OF INVENTION: NIMA INTERACTING PROTEINS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/555,912A
FILING DATE: 13-NOV-1995

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Halle, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07251/011001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 37 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
IMMEDIATE SOURCE:
CLONE: FE65/Rat
US-08-555-912A-14

Query Match 100.0%; Score 21; DB 2; Length 37;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 DTSG 4
1111
Db 12 DTSG 15

RESULT 9
US-08-348-518C-24
Sequence 24, Application US/08348518C
Patent No. 6022740
GENERAL INFORMATION:
APPLICANT: SUDOL, MARIUS
APPLICANT: PEER, BORK
APPLICANT: HENRY, CHEN
TITLE OF INVENTION: A SH2 DOMAIN ASSOCIATED PROTEIN, A
TITLE OF INVENTION: SIGNALING DOMAIN THEREIN, NUCLEIC ACIDS ENCODING THE
TITLE OF INVENTION: PROTEIN AND THE DOMAIN, AND DIAGNOSTIC AND THERAPEUTIC USES
TITLE OF INVENTION: THEREOF
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/348,518C
FILING DATE: 01-DEC-1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-101
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 37 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO

FRAGMENT TYPE: Internal
ORIGINAL SOURCE:
ORGANISM: Rat
IMMEDIATE SOURCE:
CLONE: FE65
US-08-348-518C-24

Query Match 100.0%; Score 21; DB 3; Length 37;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DTSG 4
1111
DB 12 DTSG 15

RESULT 10
US-08-476-509B-24
Sequence 24, Application US/08476509B
Patent No. 6034212
GENERAL INFORMATION:

APPLICANT: SUDOL, MARTIN
APPLICANT: PEER, BORK
APPLICANT: HENRY, CHEN
TITLE OF INVENTION: A SH2 DOMAIN ASSOCIATED PROTEIN, A
TITLE OF INVENTION: SIGNALING DOMAIN THEREIN, NUCLEIC ACIDS ENCODING THE
TITLE OF INVENTION: PROTEIN AND THE DOMAIN, AND DIAGNOSTIC AND THERAPEUTICS
TITLE OF INVENTION: THEREOF
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/476.509B
FILING DATE: 01-DEC-1994
CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-101 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521

INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 37 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
FRAGMENT TYPE: Internal
ORIGINAL SOURCE:
ORGANISM: Rat
IMMEDIATE SOURCE:
CLONE: FE65
US-08-476-509B-24

Query Match 100.0%; Score 21; DB 3; Length 37;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DTSG 4
1111
DB 12 DTSG 15

RESULT 11
US-09-188-930-322
Sequence 322, Application US/09188930A
Patent No. 6150502
GENERAL INFORMATION:

APPLICANT: Watson, James D.
APPLICANT: Strachan, Lorna
APPLICANT: Sleeman, Matthew
APPLICANT: Onrust, Rene
APPLICANT: Murison, James Greg
TITLE OF INVENTION: Compositions Isolated From Skin Cells
TITLE OF INVENTION: and Methods For Their Use
FILE REFERENCE: 11000.1011C1
CURRENT APPLICATION NUMBER: US/09/188,930A
CURRENT FILING DATE: 1998-11-09
NUMBER OF SEQ ID NOS: 348
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 322
LENGTH: 54
TYPE: PRT
ORGANISM: Mouse
US-09-188-930-322

Query Match 100.0%; Score 21; DB 4; Length 54;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DTSG 4
1111
DB 5 DTSG 8

RESULT 12
US-08-475-213-3
Sequence 3, Application US/08475213
Patent No. 5783674
GENERAL INFORMATION:

APPLICANT: Geysen, Hendrik M.
TITLE OF INVENTION: Method for the use and Synthesis of
TITLE OF INVENTION: Peptides
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/475,213
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/752,437
FILING DATE: 06-SEP-1991

PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/au90/00062
FILING DATE: 16-FEB-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU P32788/89
FILING DATE: 17-FEB-1989
ATTORNEY/AGENT INFORMATION:

NAME: Green, Grant D.
REGISTRATION NUMBER: 31259
REFERENCE/DOCKET NUMBER: 0240.002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-601-2706
TELEFAX: 510-655-3542
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 66 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHEetical: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-475-213-3

Query Match 100.0%; Score 21; DB 1; Length 66;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTSG 4
DB 18 DTSG 21

RESULT 13
US-09-276-071-2
Sequence 2, Application US/09276071
Patent No. 6207149
GENERAL INFORMATION:
APPLICANT: Fuglsang, Claus Cicone
APPLICANT: Tsuchiya, Rie
TITLE OF INVENTION: Starch Binding Domains (SBDs) For Oral Care Products
FILE REFERENCE: 5017 204-US
CURRENT APPLICATION NUMBER: US/09/276,071
CURRENT FILING DATE: 1999-03-25
NUMBER OF SEQ ID NOS: 7
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 109
TYPE: PRT
ORGANISM: Steatothermophilus
US-09-276-071-2

Query Match 100.0%; Score 21; DB 4; Length 109;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTSG 4
DB 38 DTSG 41

RESULT 14
US-08-478-039-78
Sequence 78, Application US/08478039
Patent No. 5681722
GENERAL INFORMATION:
APPLICANT: Newman, Roland A.
APPLICANT: Hanna, Nabill
APPLICANT: Raab, Ronald W.
TITLE OF INVENTION: Recombinant Antibodies for Human Therapy
NUMBER OF SEQUENCES: 114
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
STREET: 699 Prince St.
CITY: Alexandria
STATE: VA
COUNTRY: USA

ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,039
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/379,072
FILING DATE: 25-JAN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/912,292
FILING DATE: 10-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/856,281
FILING DATE: 23-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/735,064
FILING DATE: 25-JUL-1991
ATTORNEY/AGENT INFORMATION:
NAME: Teskin Esq., Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-160
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6620
TELEFAX: 703-836-2021
INFORMATION FOR SEQ ID NO: 78:
SEQUENCE CHARACTERISTICS:
LENGTH: 124 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Monkey
POSITION IN GENOME:
CHROMOSOME/SEGMENT: VH4 clone 4-16
US-08-478-039-78

Query Match 100.0%; Score 21; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTSG 4
DB 75 DTSG 78

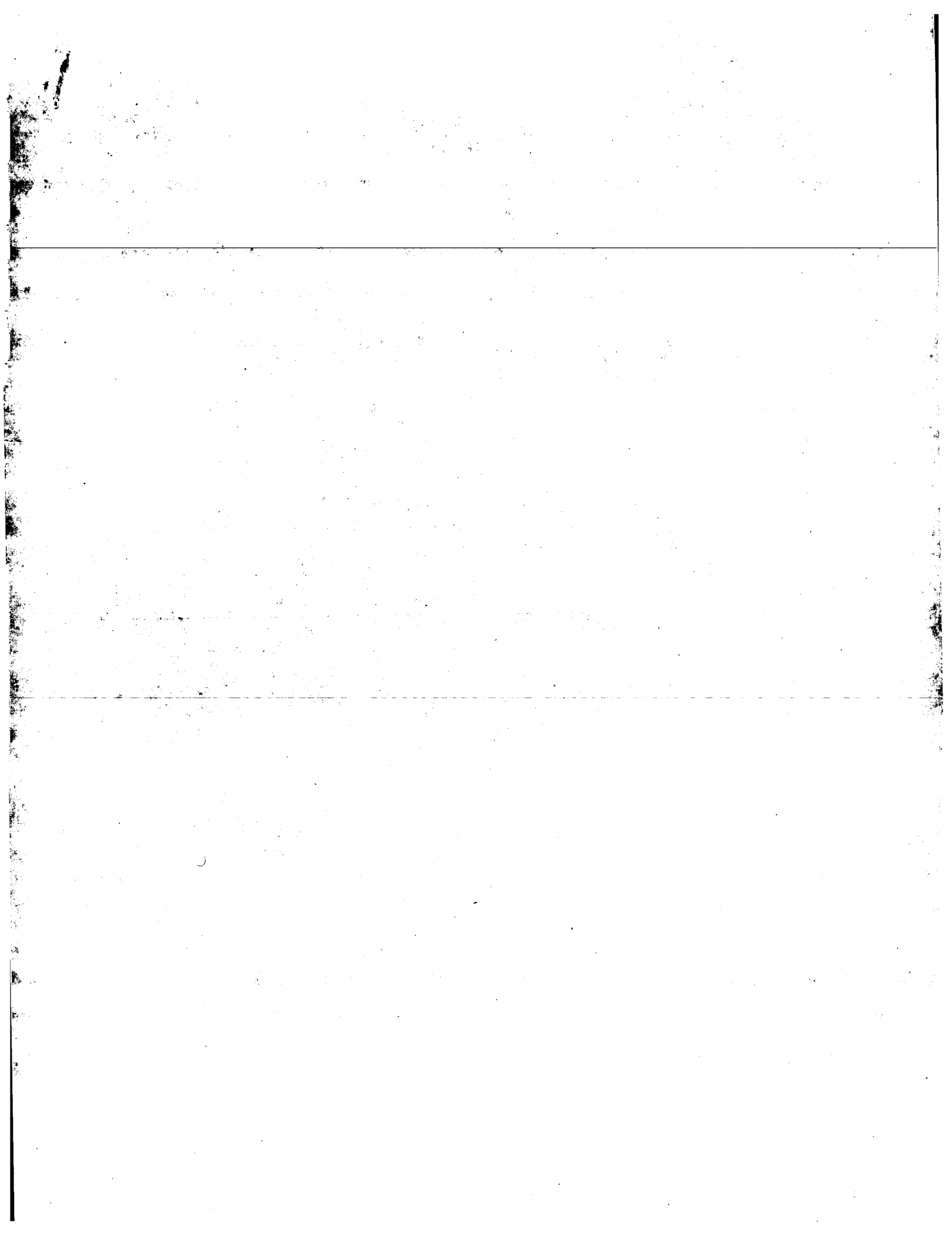
RESULT 15
US-08-476-349A-78
Sequence 78, Application US/08476349A
Patent No. 5750105
GENERAL INFORMATION:
APPLICANT: Newman, Roland A.
APPLICANT: Hanna, Nabill
APPLICANT: Raab, Ronald W.
TITLE OF INVENTION: Recombinant Antibodies for Human Therapy
NUMBER OF SEQUENCES: 114
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
STREET: 699 Prince St.
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/476,349A
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/379,072
: FILING DATE: 25-JAN-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/912,292
: FILING DATE: 10-JUL-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/856,281
: FILING DATE: 23-MAR-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/735,064
: FILING DATE: 25-JUL-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Teskin Esq., Robin L.
: REGISTRATION NUMBER: 35,030
: REFERENCE/DOCKET NUMBER: 012712-161
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 703-836-6620
: TELEFAX: 703-836-2021
: INFORMATION FOR SEQ ID NO: 78:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 124 amino acids
: TYPE: amino acid
: STRANDEDNESS: not relevant
: TOPOLOGY: not relevant
: MOLECULE TYPE: peptide
: ORIGINAL SOURCE:
: ORGANISM: Monkey
: POSITION IN GENOME:
: CHROMOSOME/SEGMENT: VH4 clone 4-16
: US-08-476-349A-78

Query Match 100.0%; Score 21; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTSG 4
1111
Db 75 DTSG 78

Search completed: September 6, 2001, 16:39:36
Job time: 125 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:43:28 : Search time 225.25 Seconds
(without alignments)
1.077 Million cell updates/sec

Title: US-09-603-713-8

Perfect score: 21

Sequence: 1 DTSC 4

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

A.Geneseq_0601.*
1: /SID58/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID58/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID58/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID58/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID58/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID58/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID58/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID58/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID58/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID58/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID58/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID58/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID58/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID58/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID58/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID58/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID58/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID58/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID58/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID58/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID58/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID58/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	4	22	AA66578
2	21	100.0	4	22	AA661340
3	21	100.0	14	21	AAV71942
4	21	100.0	15	17	AA607926
5	21	100.0	17	18	AAW15543
6	21	100.0	21	18	AAW26587
7	21	100.0	21	18	AAW14435
8	21	100.0	25	21	AAV56813
9	21	100.0	29	21	AAW21948
10	21	100.0	30	15	AAW53481
11	21	100.0	37	17	AAW97691

12	21	100.0	37	21	AA621981
13	21	100.0	38	21	AA656323
14	21	100.0	39	21	AA619101
15	21	100.0	54	21	AA657664
16	21	100.0	54	21	AA676067
17	21	100.0	54	22	AA656006
18	21	100.0	55	21	AA602277
19	21	100.0	66	21	AA600247
20	21	100.0	67	21	AAV91445
21	21	100.0	67	20	AAV02677
22	21	100.0	69	21	AA602148
23	21	100.0	73	21	AA619100
24	21	100.0	83	21	AA651439
25	21	100.0	83	21	AA651444
26	21	100.0	83	21	AA651450
27	21	100.0	83	21	AA651454
28	21	100.0	83	21	AA651457
29	21	100.0	88	21	AA600610
30	21	100.0	89	20	AAV11787
31	21	100.0	89	21	AA601562
32	21	100.0	92	20	AAV36856
33	21	100.0	93	21	AA612144
34	21	100.0	94	21	AA612143
35	21	100.0	96	21	AA612142
36	21	100.0	97	19	AAW71021
37	21	100.0	97	19	AAW70950
38	21	100.0	99	21	AAV44882
39	21	100.0	105	20	AAW89029
40	21	100.0	105	21	AA602854
41	21	100.0	106	21	AA642474
42	21	100.0	109	16	AA669091
43	21	100.0	109	18	AAW08740
44	21	100.0	109	19	AAW57249
45	21	100.0	109	20	AAW93897

ALIGNMENTS

RESULT 1	
AA66578	standard; Peptide: 4 AA.
AC	AA66578;
DT	12-APR-2001 (first entry)
DE	Signature aspartic protease motif.
KW	Memapsin 2; neotropic; neuroprotective; amyloid precursor protein; APP; memapsin 2 inhibitor; Alzheimer's disease; aspartic protease.
OS	Unidentified.
XX	
PN	WO200100665-A2.
PD	04-JAN-2001.
PF	27-JUN-2000; 2000WO-US17742.
PR	28-JUN-1999; 99US-0141363.
PR	30-NOV-1999; 99US-0168060.
PR	25-JAN-2000; 2000US-0177836.
PR	27-JAN-2000; 2000US-0178368.
PR	08-JUN-2000; 2000US-0210292.
XX	
PA	(OKLA-) OKLAHOMA MEDICAL RES FOUND.
PA	(UNIT) UNIT ILLINOIS FOUND.
PI	Tang JUN, Hong L, Ghosh AK.
PI	
DR	WPI; 2001-137933/14.
XX	

FE65/rat peptide c
Human secreted pro
Zea mays protein f
Arabidopsis thalia
Murine skin cell t
Skin cell protein,
Human secreted pro
Human secreted pro
Human secreted pro
Human secreted pro
Human secreted pro
Zea mays protein f
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Human secreted pro
Human secreted pro
Human 5' EST secre
Amino acid sequenc
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Mus musculus Cgamm
Light chain of mon
Polyptide transcrip
Human secreted pro
Human ORFX ORF238
Anti-HIV Fab
Human anti-HIV Fab
Amino acid sequenc
B. stearothermophi

PT Novel memapsin 2 inhibitors which bind to active site of memapsin 2
 PT having 2 catalytic aspartic residues and substrate binding cleft, used
 PT to treat Alzheimer's disease by blocking amyloid precursor protein
 PT cleavage -
 XX
 PS Disclosure; Page 14; 86pp; English.
 CC
 CC The present sequence is given in a specification relating to an inhibitor
 CC of catalytically active memapsin 2. The inhibitor binds to an inhibitory
 CC active site, which is defined by the presence of two catalytic aspartic
 CC residues and a substrate binding cleft. The inhibitor is useful for
 CC the treatment and diagnosis of Alzheimer's disease. It is useful in
 CC screens for individuals with a genetic predisposition to Alzheimer's
 CC disease. The inhibitor is useful as a reagent for specifically binding to
 CC memapsin 2 or memapsin 2 analogues and for aiding in memapsin 2
 CC isolation, purification and characterisation.
 CC
 SO Sequence 4 AA;

Query Match 100.0%; Score 21; DB 22; Length 4;
 Best Local Similarity 100.0%; Pred. No. 3.4e+05;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTSG 4
 ||||
 Db 1 dtsg 4

RESULT 2

ID AAB61340 standard; peptide; 4 AA.

AC AAB61340;

DT 02-APR-2001 (first entry)

DE Aspartic protease sequence motif.

XX Memapsin 2; catalyst; Alzheimer's.

OS Unidentified.

PN WO200100663-A2.

PD 04-JAN-2001.

PF 27-JUN-2000; 2000WO-US17661.

PR 28-JUN-1999; 99US-0141363.

PR 30-NOV-1999; 99US-0168060.

PR 25-JAN-2000; 2000US-0177836.

PR 27-JAN-2000; 2000US-0178368.

PR 08-JUN-2000; 2000US-0210292.

XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.

PI Tang JN, Lin X, Koelsch G;

DR WPI; 2001-102885/11.

XX Purified recombinant catalytically active memapsin 2, used to screen

PT inhibitors of it, which are used to treat and prevent Alzheimer's

PT disease -

XX Disclosure; Page 14; 86pp; English.

CC The present invention relates to a purified recombinant

CC catalytically active memapsin 2. The invention may be used for

CC isolating inhibitors which are used to treat or prevent

CC Alzheimer's disease. The invention may also be used to screen

CC for individuals more genetically prone to develop Alzheimer's

CC disease.

XX
 SO Sequence 4 AA;

Query Match 100.0%; Score 21; DB 22; Length 4;
 Best Local Similarity 100.0%; Pred. No. 3.4e+05;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTSG 4
 ||||
 Db 1 dtsg 4

RESULT 3

ID AAY71942 standard; peptide; 14 AA.

AC AAY71942;

DT 26-MAR-2001 (first entry)

DE CDR3 of heavy chain variable region (VH) of human CS37 antibody.

XX Human; heavy chain variable region; VH; CS37 antibody; cytostatic;

KW ophthalmological; immunomodulatory; antiinflammatory; antileukemic;

KW antiaesthetic; transforming growth factor-beta 1; TGF-beta 1; CDR3;

KW complementarity determining region 3; treatment; glomerulonephritis;

KW keloid; hypertrophic scarring; proliferative vitreoretinopathy; cataract;

KW glaucoma drainage surgery; corneal injury; immune system; asthma; tumour;

KW inflammatory response; angiogenesis; metastasis; leukaemia; sarcoma;

KW fibrosis; rheumatoid arthritis; hepatitis B; hepatitis C; AIDS;

KW acquired immune deficiency syndrome; extracellular matrix deposition.

XX Homo sapiens.

OS WO20006631-A1.

PN 09-NOV-2000.

PD 02-MAY-2000; 2000WO-GB01679.

PR 30-APR-1999; 99US-0131983.

PR (CAMP-) CAMBRIDGE ANTIBODY TECHNOLOGY.

XX Thompson JE, Leonard SN, Wilton AJ, Braddock PSH, Du Fou SL;

PI McCallerty JG, Conroy LA, Tempest PR;

XX WPI; 2000-687531/67.

DR Antibodies that specifically bind to transforming growth factor-beta1,

XX useful for treating e.g. cancers and ophthalmological disorder such as

XX cataracts and proliferative vitreoretinopathy -

XX Example 1; Page 5; 84pp; English.

PS The present invention relates to specific binding members

XX which are capable of binding to transforming growth factor-beta 1

CC (TGF-beta 1). The invention relates specifically to the antibodies

CC which include the complementarity determining region 3 (CDR3) of

CC the heavy chain variable domain (VH) of the S15 antibody and J182

CC antibody. These antibodies have strong neutralising activity for

CC TGF-beta 1 and are used for the treatment of a human or an animal

CC suffering from a condition associated with TGF-beta 1 expression and

CC extracellular matrix deposition such as glomerulonephritis, keloid and

CC hypertrophic scarring, proliferative vitreoretinopathy, glaucoma drainage

CC surgery, corneal injury and cataracts. They are used to treat asthma,

CC fibrosis, rheumatoid arthritis and tumours associated with angiogenesis

CC and/or metastasis such as breast, prostate, ovarian, stomach colorectal,

CC skin, lung, cervical and/or bladder tumours, leukaemia and/or sarcomas.

CC They are also used to modulate the immune system and inflammatory

CC responses and improve the immune response to infections such as

CC hepatitis B, hepatitis C and AIDS.

CC The present sequence is a complementarity determining region 3 (CDR3)
 CC of heavy chain variable region (VH) of human CS37 antibody.
 XX
 SO Sequence 14 AA:

Query Match 100.0%; Score 21; DB 21; Length 14;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTSG 4
 ||||
 Db 8 dtsg 11

RESULT 4

AAW07926 standard; peptide; 15 AA.

AC AAW07926;

DT 31-JAN-1997 (first entry)

DE gp41 peptide MN-8.

KW HIV; gp120; HIV-IIIB strain; HIV-1 transmission; foetal transmission;
 neutralising antibody; passive immunisation; anti-idiotypic antibody;
 KM gp41; vaccine; active immunotherapy.

OS Human immunodeficiency virus type 1.

PN US5556744-A.

PD 17-SEP-1996.

XX 24-MAR-1994; 94US-0218025.

XX 24-MAR-1994; 94US-0218025.

PR 29-MAY-1992; 92US-0891451.

PA (UYPE-) UNIV PENNSYLVANIA.

PI (WIST-) WISTAR INST ANATOMY & BIOLOGY.

PI Ugen KE, Weiner DB, Williams WV;

DR MPI: 1996-432980/43.

PT Determining the likelihood of maternal transmission of HIV-1 to
 foetus - by measuring maternal reactivity with specific gp120 and
 PT gp41 derived peptide(s), also used for diagnosing HIV in infants

PS Claim 1: Column 115; 63pp; English.

CC AAW07918-W07928 represent HIV gp41 peptides that can be used in the
 CC method of the invention. These sequences are derived from the HIV-IIIB
 CC strain deposited as ATCC CRL 8543. The method of the invention is for
 CC determining whether or not a mother will transmit HIV-1 to a foetus.
 CC The method comprises incubating a sample from the HIV-infected mother,
 CC with a collection of HIV peptides. The HIV peptides includes at least
 CC one of these sequences, and at least one HIV gp120 derived peptide (see
 CC AAW07909-W07917). The number of peptides that react with the sample is
 CC determined, and this number is compared with a standard that shows
 CC non-transmissible HIV sample is indicated if the test sample reacts with
 CC twice as many peptides as the standard. The method detects the presence
 CC of neutralising antibodies that protect against mother to infant
 CC transmission of HIV. These sequences can also be used in vaccines to
 CC protect against transmission. Antibodies against these sequences can be
 CC used for passive immunisation, and to generate anti-idiotypic antibodies
 CC for use in vaccines or active immunotherapy.

SO Sequence 15 AA:

Query Match 100.0%; Score 21; DB 17; Length 15;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTSG 4
 ||||
 Db 1 dtsg 4

RESULT 5

AAW15543 standard; peptide; 17 AA.

AC AAW15543;

DT 27-NOV-1997 (first entry)

DE Anti-TGF beta-1 scfv antibody 1-B2 VH domain CDR3 mutant 27-C1.

KW Transforming growth factor beta-1; TGF-beta-1; human; therapy;
 KW antibody engineering; scfv; phage display; lung fibrosis;
 KW arterial injury; proliferative retinopathy; retinal detachment;
 KW adult respiratory distress syndrome; liver cirrhosis;
 KW post myocardial infarction; post-angioplasty restenosis;
 KW scleroderma; vascular disease; cataract; glaucoma; scarring;
 KW glomerulonephritis; osteoporosis; immune disease; inflammation;
 KW rheumatoid arthritis; macrophage deficiency disease;
 KW macrophage pathogen infection; complementarity determining region;
 KW CDR.

OS Chimeric Homo sapiens;

OS Chimeric synthetic.

PH Key Location/Qualifiers

FT Misc-difference 11 /note="Ser in 1B2 CDR3"

PN GB2305921-A.

PD 23-APR-1997.

XX 07-OCT-1996; 96GB-0020920.

XX 19-JAN-1996; 96GB-0001081.

PR 06-OCT-1995; 95GB-0020486.

PA (CAMP-) CAMBRIDGE ANTIBODY TECHNOLOGY.

PI Bacon L, Green JA, Jackson RH, Johnson KS, Pope AR;

PI Tempest PR, Thompson JE, Vaughan TJ, Williams AJ;

PI Wilton AJ;

DR MPI: 1997-215360/20.

PT Agent contg. antigen-binding domain of human antibody to
 PT transforming growth factor beta 1 or 2 - and nucleic acid encoding
 PT it, used to neutralise effects of TGF, e.g. for control of fibrosis,
 PT immune and inflammatory disease

PS Claim 20: Fig 3; 184pp; English.

CC Variants (AAW15542-58) of the complementarity determining region 3
 CC (CDR3) of human scfv antibody 1B2 VH domain were produced by
 CC splicing mutagenesis. 1B2 is specific for transforming growth
 CC factor (TGF) beta-1. Novel human antibodies that contain these
 CC CDR3 variants have a preference for TGF beta-1 over TGF beta-2. The
 CC antigen-binding domains of human antibodies (see AAW15522-40) to TGF
 CC beta-1 and/or beta-2 can be used to counter the adverse effects of
 CC TGF beta, such as (i) promotion of fibrosis (in dermal, ocular or
 CC keloid scarring, lung fibrosis, arterial injury, proliferative
 CC retinopathy, retinal detachment, adult respiratory distress syndrome,
 CC liver cirrhosis, post myocardial infarction, post-angioplasty

CC restenosis, scleroderma, vascular disorders, cataract, glaucoma, or
CC esp. neural scarring and glomerulonephritis, also (not claimed)
CC osteoporosis), or (ii) immune and inflammatory diseases (rheumatoid
CC arthritis, macrophage deficiency diseases or macrophage pathogen
CC infection).
XX
SQ Sequence 17 AA;

Query Match 100.0%; Score 21; DB 18; Length 17;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DTSG 4
|
|
|
|
DB 10 dtsg 13

RESULT 6
AAW26587
ID AAW26587 standard; peptide: 21 AA.

XX AAW26587;
AC
XX
XX 21-JAN-1998 (first entry)
DT
XX
XX Hemidesmosome formation inducing protein 154 kDa subunit peptide.
DE
XX Hemidesmosome: laminin; epithelial cell; cell attachment; adhesion;
KW bladder carcinoma; 804G; NBT-II; rat; dental implant.
KM
XX
OS Rattus sp.
XX
XX US5658789-A.
PN
XX 19-AUG-1997.
PD
XX 12-NOV-1993; 93US-0151134.
PF
XX 19-MAY-1995; 95US-0445135.
PR
XX 12-NOV-1993; 93US-0151134.
XX
XX (DESM-) DESMOS INC.
PA
XX Hormla M, Quaranta V;
PI
XX WPI; 1997-424242/39.
DR
XX

XX Soluble protein that induces hemidesmosome formation in epithelial
PT cells - useful for stimulating epithelial cell attachment, e.g. to
PT dental implants or teeth
XX
XX Example 11; Column 31-32; 20pp; English.

XX This peptide comprises an internal peptide of the 154 kDa alpha
CC chain (see also AAW26584) of an isolated soluble protein that induces
CC hemidesmosome formation in epithelial cells normally unable to form
CC hemidesmosomes. The soluble protein is obtainable from 804G rat
CC bladder carcinoma cells or NBT-II rat bladder carcinoma cells. It
CC can be used to induce hemidesmosome formation in epithelial cells
CC and to facilitate their growth.
XX
SQ Sequence 21 AA;

Query Match 100.0%; Score 21; DB 18; Length 21;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DTSG 4
|
|
|
|
DB 17 dtsg 20

RESULT 7
AAW14435
ID AAW14435 standard; Protein: 21 AA.
XX
XX AAW14435;
AC
XX
XX 14-MAY-1997 (first entry)
DT
XX
XX H6 protein epitope.
DE
XX
XX cotton; fibre-specific; strength; transgenic plant; anthesis;
KW developmentally regulated; E6; H6; antisense; sense; epitope.
KM
XX
XX Synthetic.
OS
XX
XX US5597718-A.
PN
XX
XX 28-JAN-1997.
PD
XX
XX 04-OCT-1988; 88US-0253243.
PE
XX
XX 20-SEP-1995; 95US-0530797.
PR
XX 04-OCT-1988; 88US-0253243.
PR 21-NOV-1990; 90US-0617239.
PR 18-OCT-1993; 93US-0138814.
XX
XX (CETU) AGRACETUS.
PA
XX
XX Brill WJ, John ME, Umbeck PF;
PI
XX WPI; 1997-108326/10.
DR
XX

XX Prod. of transgenic cotton plants - by transformation with the H6
PT coding sequence or E6 anti-sense sequence, produces fibre of altered
PT strength
XX
XX Example 9; Column 15; 33pp; English.
PS
XX
XX AAW14434-35 are peptide epitopes of the E6 and H6 proteins respectively.
CC Antibodies against these peptides were raised and used to characterise
CC the location of the proteins in fibre cells. Cotton fibre-specific cDNA
CC clones (AAW162609-24) can be used to identify genomic clones by
CC differential cDNA library screenings. Coding sequences from these
CC isolated genes are used in sense or antisense orientation to alter the
CC fibre characteristics, e.g. strength, of transgenic fibre-producing
CC plants.
CC
XX
SQ Sequence 21 AA;

Query Match 100.0%; Score 21; DB 18; Length 21;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DTSG 4
|
|
|
|
DB 16 dtsg 19

RESULT 8
AAW56813
ID AAW56813 standard; peptide: 25 AA.
XX
XX AAW56813;
AC
XX
XX 31-MAR-2000 (first entry)
DT
XX
XX Neuronal glutamate receptor GluR2 peptide fragment.
DE
XX
XX Rasmussen's encephalitis; autoantibody; glutamate receptor; GluR3;
KW neuronal; GluR2.
XX

OS Synthetic.
 XX US6010854-A.
 PN
 XX 04-JAN-2000.
 PD
 XX 03-JUL-1997; 97US-0887769.
 XX
 XX 28-NOV-1994; 94US-0345527.
 XX
 XX (UTAH) UNIV UTAH RES FOUND.
 PA
 XX Gahring LC, Tryman RE, Rogers SW;
 PI
 XX WPI; 2000-105494/09.
 DR
 XX Screening assay for Rasmussen's encephalitis useful for diagnosis and
 XX for monitoring treatment -
 PT
 XX Disclosure; Fig 2; 19pp; English.
 PS
 XX The invention provides a screening assay for Rasmussen's encephalitis
 XX that comprises detecting autoantibodies specific for one of two defined
 CC amino acid sequences (AAY56808-809) of glutamate receptor GluR3. The
 CC presence of such autoantibodies signifies that the patient has or is
 CC predisposed to Rasmussen's encephalitis. The method is used for
 CC diagnosis and monitoring of Rasmussen's encephalitis. The present
 CC sequence represents a peptide from neuronal glutamate receptor GluR2
 CC used for homology studies with the GluR3 peptide.
 XX
 XX Sequence 25 AA;
 SQ

Query Match 100.0%; Score 21; DB 21; Length 25;
 Best Local Similarity 100.0%; Pred. No. 69;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTSG 4
 Db 18 dtsg 21
 ||||

RESULT 9
 AAB21948
 ID AAB21948 standard; peptide; 29 AA.
 XX
 AC AAB21948;
 XX
 DT 02-JAN-2001 (first entry)
 XX
 DE FE65/rat peptide containing a WW-domain #1.
 XX
 KW WW-domain; protein-protein interaction; cell growth regulation;
 KW protein degradation regulation; Alzheimer's; Dementia pugilistica;
 KW Down's syndrome; Parkinson's disease; Pick's; neurodegenerative;
 KW microtubule assembly; tau; hyperplasia; neoplasia; malignancy;
 KW psoriasis; retinosis; atherosclerosis; leukaemia; lymphoma; papilloma;
 KW pulmonary fibrosis; rheumatoid arthritis; multiple sclerosis;
 KW muscular dystrophy; rat.
 XX
 OS Rattus sp.
 XX
 PN WO200048621-A2.
 XX
 XX 24-AUG-2000.
 PD
 XX 18-FEB-2000; 2000WO-US04278.
 PF
 XX 18-FEB-1999; 99US-0252404.
 XX
 XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 PA
 XX Lu KP, Zhou XZ;
 PI

XX WPI; 2000-594014/56.
 XX
 PT Mediating protein-protein interactions, useful for regulating cell
 PT growth and for treating neurodegenerative disorders, comprises
 PT modulating binding of WW domain containing polypeptide with
 PT phosphorylated ligand -
 XX
 XX Disclosure; Fig 2; 82pp; English.
 XX
 CC The present invention relates to a method for mediating protein-protein
 CC interaction, which comprises modulating the binding of a WW-domain
 CC containing peptide with a phosphorylated ligand e.g. tau. WW-domains are
 CC highly conserved regions of approximately 40 amino acid residues with two
 CC invariant tryptophans (W) in a triple stranded beta-sheet. The present
 CC sequence is one such WW-domain. When a WW-domain containing peptide is
 CC phosphorylated at serine or threonine residues, dephosphorylation of
 CC ligands bound to the peptide is inhibited. The present peptide may be
 CC useful for mediating protein-protein interaction, regulating cell growth,
 CC regulating protein degradation, restoring the function of tau to bind
 CC microtubules and promote or restore microtubule assembly in
 CC neurodegenerative diseases e.g. Alzheimer's, Dementia pugilistica, Down's
 CC syndrome, Parkinson's disease, Pick's disease, multiple sclerosis,
 CC muscular dystrophy, Corticobasal degeneration, Frontotemporal dementias,
 CC Myotonic dystrophy, Niemann-Pick disease, prion disease with tangles,
 CC progressive supranuclear palsy and subacute sclerosing panencephalitis.
 CC In addition, inhibitors or stimulators of interactions between WW-domains
 CC and ligands of the present invention can be used to treat hyperplastic
 CC and neoplastic disorders e.g. all forms of malignancies, psoriasis,
 CC retinosis, atherosclerosis resulting from plaque formation, leukaemias,
 CC benign tumour growth, lymphomas, papillomas, pulmonary fibrosis and
 CC rheumatoid arthritis.
 XX
 XX Sequence 29 AA;
 SQ

Query Match 100.0%; Score 21; DB 21; Length 29;
 Best Local Similarity 100.0%; Pred. No. 81;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTSG 4
 Db 10 dtsg 13
 ||||

RESULT 10
 AAR53481
 ID AAR53481 standard; peptide; 30 AA.
 XX
 AC AAR53481;
 XX
 DT 01-DEC-1994 (first entry)
 XX
 DE CD44 peptide CD44-8.
 XX
 XX Cell adhesion molecule; CD44; antiinflammatory; rheumatoid;
 KW arthritis; tumor cell metastasis; autoimmune disease;
 KW immunosuppressive.
 KW
 XX Homo sapiens.
 OS
 XX
 PN WO9409811-A.
 XX
 XX 11-MAY-1994.
 PD
 XX 29-OCT-1993; 93WO-US10412.
 PF
 XX 30-OCT-1992; 92US-0973339.
 XX
 XX (UYDU-) UNIV DUKE.
 PA
 XX Hale LP, Haynes BF, Liao H, Patton KL, Telen MJ;
 XX
 XX

DR WPI; 1994-167121/20.
 XX Use of CD44 protein and new peptide derivs - for developing prods
 PT for inflammation, immune-mediated tissue damage and tumour cell
 PT metastasis
 XX
 XX Claim 4; Page 14; 83pp; English.
 PS
 CC The peptide can be used for treating inflammation and immune-
 CC mediated tissue damage such as occurs in the course of autoimmune
 CC diseases, e.g. rheumatoid arthritis. It can also be used for
 CC determining metastatic potential or to treat or prevent tumor cell
 CC metastasis. This peptide corresponds to AA 150-161, 170-177 and
 CC 189-198 of the CD44
 CC protein sequence.
 XX
 XX Sequence 30 AA;
 SQ

Query Match 100.0%; Score 21; DB 15; Length 30;
 Best Local Similarity 100.0%; Pred. NO. 84;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 DTSG 4
 Db 20 dtsg 23
 ||||

RESULT 11
 AAR97691
 ID AAR97691 standard; Protein; 37 AA.
 XX
 XX AAR97691;
 AC
 XX
 XX 30-AUG-1996 (first entry)
 DT
 XX
 XX Rat FE65 WW domain-1.
 DE
 XX
 XX WW domain; signal transduction; diagnosis; gene therapy;
 KW Yes proto-oncogene associated protein; YAP; FE65.
 KW
 XX
 XX Rattus sp.
 OS
 XX
 XX WO9617061-A1.
 FN
 XX
 XX 06-JUN-1996.
 PD
 XX
 XX 30-NOV-1995; 95WO-US15512.
 PF
 XX
 XX 07-JUN-1995; 95US-0476509.
 PR
 XX
 XX 01-DEC-1994; 94US-0348518.
 PR
 XX
 XX (DELB-) DELBRUECK CENT MOLECULAR MEDICINE MAX.
 PA (UVRQ) UNIV ROCKEFELLER.
 PA
 XX
 XX Bork P, Chen H, Sudol M;
 PI
 XX
 XX WPI; 1996-286829/29.
 DR
 XX
 XX DNA encoding Yes proto-oncogene associated protein - used to
 PT modulate intracellular signal transduction e.g. for treatment of
 PT muscular dystrophy
 PT
 XX
 XX Claim 10; Fig 12; 126pp; English.
 PS
 XX
 XX WW domains (AAR97673-92) were identified in a number of proteins
 CC involved in signalling or regulatory functions. The WW domain was
 CC initially identified in the Yes proto-oncogene associated proteins
 CC (YAPs) of chicken, human and mouse (see also AAR97669-70 and AAR97672).
 CC A consensus sequence is given in AAR97671. Rat FE65 is a
 CC transcription factor activator expressed preferentially in liver.
 CC It contains a WW domain (AAR97691) that can be expressed in bacterial,
 CC yeast, insect or mammalian cells, and used to identify WW domain

CC ligands. WW domains can be introduced into cells, either directly or
 CC by gene therapy, to increase the level of signal transduction.
 XX
 XX Sequence 37 AA;
 SQ

Query Match 100.0%; Score 21; DB 17; Length 37;
 Best Local Similarity 100.0%; Pred. NO. 1e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 DTSG 4
 Db 12 dtsg 15
 ||||

RESULT 12
 AAB21981
 ID AAB21981 standard; peptide; 37 AA.
 XX
 XX AAB21981;
 AC
 XX
 XX 02-JAN-2001 (first entry)
 DT
 XX
 XX FE65/rat peptide containing a WW-domain #2.
 DE
 XX
 XX WW-domain; protein-protein interaction; cell growth regulation;
 KW protein degradation regulation; Alzheimer's; Dementia pugilistica;
 KW Down's syndrome; Parkinson's disease; Pick's; neurodegenerative;
 KW microtubule assembly; tau; hyperplasia; neoplasia; malignancy;
 KW psoriasis; retinosis; atherosclerosis; leukaemia; lymphoma; papilloma;
 KW pulmonary fibrosis; rheumatoid arthritis; multiple sclerosis;
 KW muscular dystrophy; rat.
 KW
 XX
 XX Rattus sp.
 OS
 XX
 XX WO200048621-A2.
 PN
 XX
 XX 24-AUG-2000.
 PD
 XX
 XX 18-FEB-2000; 2000WO-US04278.
 PF
 XX
 XX 18-FEB-1999; 99US-0252404.
 PR
 XX
 XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 PA
 XX
 XX Lu KP, Zhou XZ;
 PI
 XX
 XX WPI; 2000-594014/56.
 DR
 XX
 XX Mediating protein-protein interactions, useful for regulating cell
 PT growth and for treating neurodegenerative disorders, comprises
 PT modulating binding of WW domain containing polypeptide with
 PT phosphorylated ligand -
 PT
 XX
 XX Disclosure; Fig 6; 82pp; English.
 PS
 XX
 XX The present invention relates to a method for mediating protein-protein
 CC interaction, which comprises modulating the binding of a WW-domain
 CC containing peptide with a phosphorylated ligand e.g. tau. WW-domains are
 CC highly conserved regions of approximately 40 amino acid residues with two
 CC invariant tryptophans (W) in a triple stranded beta-sheet. The present
 CC sequence is one such WW-domain. When a WW-domain containing peptide is
 CC phosphorylated at serine or threonine residues, dephosphorylation of
 CC ligands bound to the peptide is inhibited. The present peptide may be
 CC useful for mediating protein-protein interaction, regulating cell growth,
 CC regulating protein degradation, restoring the function of tau to bind
 CC microtubules and promote or restore microtubule assembly in
 CC neurodegenerative diseases e.g. Alzheimer's, Dementia pugilistica, Down's
 CC syndrome, Parkinson's disease, Pick's disease, multiple sclerosis,
 CC muscular dystrophy, Corticobasal degeneration, Frontotemporal dementias,
 CC Myotonic dystrophy, Niemann-Pick disease, prion disease with tangles,
 CC progressive supranuclear palsy and subacute sclerosing panencephalitis.
 CC In addition, inhibitors or stimulators of interactions between WW-domains

CC and ligands of the present invention can be used to treat hyperplastic
 CC and neoplastic disorders e.g. all forms of malignancies, psoriasis,
 CC retinosis, atherosclerosis resulting from plaque formation, leukaemias,
 CC benign tumour growth, lymphomas, papillomas, pulmonary fibrosis and
 CC rheumatoid arthritis.

XX Sequence 37 AA;

Query Match 100.0%; Score 21; DB 21; Length 37;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 DTSG 4
 Db 12 dtsg 15

RESULT 13
 AAB56323
 ID AAB56323 standard; Protein: 38 AA.
 XX AC AAB56323;
 XX DT 13-MAR-2001 (first entry)
 XX DE Human secreted protein sequence encoded by gene 106 SEQ ID NO:417.
 XX KW Human: secreted protein; diagnosis; immunosuppressive; antiarthritic;
 KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
 KW cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;
 KW fungicide; ophthalmological; gene therapy; pathological condition;
 KW autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;
 KW neoplasm; cardiovascular disorder; cardiac arrest; cerebral ischaemia;
 KW cerebrovascular disorder; angiogenesis; nervous system disorder;
 KW Alzheimer's disease; infection; ocular disorder; corneal infection;
 KW wound healing; skin aging; food additive; preservative.

XX OS Homo sapiens.
 XX PN WO200070042-A1.
 XX PD 23-NOV-2000.
 XX PF 11-MAY-2000; 2000WO-US12788.
 XX PR 13-MAY-1999; 99US-0134068.
 XX PA (HUMA-) HUMAN GENOME SCI INC.
 XX PI Rosen CA, Ruben SM, Moore PA, Young PE, Komatsoulis GA, Birse CE;
 PI Duan RD, Florence KA, Soppet DR;
 XX WPI: 2000-679828/66.
 XX PT Isolated nucleic acid molecule encoding a human secreted protein is
 PT used in preventing, treating or ameliorating a medical condition -
 XX Disclosure; Page 1044; 1065pp; English.
 XX CC The polynucleotide sequences given in AAC99818 to AAC99977 encode the
 CC human secreted proteins based in AAB56077 to AAB56362. Human secreted
 CC proteins have activities based on the tissues and cells the genes are
 CC expressed in. Examples of activities include: immunosuppressive;
 CC antirheumatic; antiproliferative; cytostatic; cardiant;
 CC vasotropic; cerebroprotective; neurotropic; neuroprotective; antibacterial;
 CC virucide; fungicide; and ophthalmological. The human secreted
 CC polynucleotides and proteins can be used to prevent, treat or ameliorate
 CC a medical condition in e.g. humans, mice, rabbits, goats, horses, cats,
 CC dogs, chickens or sheep. They are also used in diagnosing a pathological
 CC condition or susceptibility to a pathological condition. Disorders which
 CC are diagnosed or treated include autoimmune diseases e.g. rheumatoid
 CC arthritis, hyperproliferative disorders e.g. neoplasms of the breast or

CC liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular
 CC disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders
 CC e.g. Alzheimer's disease, infections caused by bacteria, viruses and
 CC fungi and ocular disorders e.g. corneal infection. The proteins can also
 CC be used to aid wound healing and epithelial cell proliferation, to
 CC prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The proteins can also be used as a
 CC food additive or preservative to increase or decrease storage
 CC capabilities. AAC9809 to AAC9817 and AAB56076 represent sequences used
 CC in the exemplification of the present invention.

XX Sequence 38 AA;

Query Match 100.0%; Score 21; DB 21; Length 38;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 DTSG 4
 Db 3 dtsg 6

RESULT 14
 AAG19101
 ID AAG19101 standard; Protein: 39 AA.
 XX AC AAG19101;
 XX DT 17-OCT-2000 (first entry)
 XX DE Zea mays protein fragment SEQ ID NO: 20773.
 XX KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence; corn.
 XX OS Zea mays subsp. mays.
 XX PN EP1033405-A2.
 XX PD 06-SEP-2000.
 XX PF 25-FEB-2000; 2000EP-0301439.
 XX PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 04-MAY-1999; 99US-0132407.
 PR 05-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.

PR	18-MAY-1999;	99US-0134768.	PR	27-JUL-1999;	99US-0145919.
PR	19-MAY-1999;	99US-0134941.	PR	28-JUL-1999;	99US-0145951.
PR	20-MAY-1999;	99US-0135124.	PR	02-AUG-1999;	99US-0146386.
PR	21-MAY-1999;	99US-0135353.	PR	02-AUG-1999;	99US-0146388.
PR	24-MAY-1999;	99US-0135629.	PR	03-AUG-1999;	99US-0146389.
PR	25-MAY-1999;	99US-0136021.	PR	04-AUG-1999;	99US-0147038.
PR	27-MAY-1999;	99US-0136392.	PR	04-AUG-1999;	99US-0147204.
PR	28-MAY-1999;	99US-0136782.	PR	05-AUG-1999;	99US-0147302.
PR	01-JUN-1999;	99US-0137222.	PR	05-AUG-1999;	99US-0147192.
PR	03-JUN-1999;	99US-0137528.	PR	06-AUG-1999;	99US-0147260.
PR	04-JUN-1999;	99US-0137502.	PR	06-AUG-1999;	99US-0147303.
PR	07-JUN-1999;	99US-0137724.	PR	09-AUG-1999;	99US-0147416.
PR	08-JUN-1999;	99US-0138094.	PR	09-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	99US-0138540.	PR	10-AUG-1999;	99US-0147935.
PR	10-JUN-1999;	99US-0138847.	PR	11-AUG-1999;	99US-0148371.
PR	14-JUN-1999;	99US-0139119.	PR	11-AUG-1999;	99US-0148319.
PR	16-JUN-1999;	99US-0139452.	PR	12-AUG-1999;	99US-0148341.
PR	16-JUN-1999;	99US-0139453.	PR	13-AUG-1999;	99US-0148565.
PR	17-JUN-1999;	99US-0139454.	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139455.	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139456.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139457.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139460.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139462.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139463.	PR	26-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139750.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151066.
PR	21-JUN-1999;	99US-0139889.	PR	27-AUG-1999;	99US-0151080.
PR	23-JUN-1999;	99US-0140353.	PR	30-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140354.	PR	31-AUG-1999;	99US-0151438.
PR	24-JUN-1999;	99US-0140695.	PR	01-SEP-1999;	99US-0151930.
PR	28-JUN-1999;	99US-0140823.	PR	07-SEP-1999;	99US-0152363.
PR	29-JUN-1999;	99US-0140991.	PR	10-SEP-1999;	99US-0153070.
PR	30-JUN-1999;	99US-0141287.	PR	13-SEP-1999;	99US-0153758.
PR	01-JUL-1999;	99US-0141842.	PR	15-SEP-1999;	99US-0154018.
PR	01-JUL-1999;	99US-0142154.	PR	16-SEP-1999;	99US-0154039.
PR	02-JUL-1999;	99US-0142055.	PR	20-SEP-1999;	99US-0154779.
PR	06-JUL-1999;	99US-0142390.	PR	22-SEP-1999;	99US-0155139.
PR	08-JUL-1999;	99US-0142803.	PR	23-SEP-1999;	99US-0155486.
PR	09-JUL-1999;	99US-0142920.	PR	24-SEP-1999;	99US-0155659.
PR	12-JUL-1999;	99US-0142977.	PR	28-SEP-1999;	99US-0156458.
PR	13-JUL-1999;	99US-0143542.	PR	29-SEP-1999;	99US-0156596.
PR	14-JUL-1999;	99US-0143624.	PR	05-OCT-1999;	99US-0157117.
PR	15-JUL-1999;	99US-0144005.	PR	06-OCT-1999;	99US-0157753.
PR	16-JUL-1999;	99US-0144085.	PR	07-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144086.	PR	12-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144325.	PR	13-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144332.	PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144333.	PR	14-OCT-1999;	99US-0159330.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159331.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159637.
PR	20-JUL-1999;	99US-0144352.	PR	18-OCT-1999;	99US-0159638.
PR	20-JUL-1999;	99US-0144632.	PR	21-OCT-1999;	99US-0159584.
PR	20-JUL-1999;	99US-0144632.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0144814.	PR	21-OCT-1999;	99US-0160767.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160768.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160815.
PR	22-JUL-1999;	99US-0145089.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145192.	PR	22-OCT-1999;	99US-0160981.
PR	23-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0160989.
PR	23-JUL-1999;	99US-0145218.	PR	25-OCT-1999;	99US-0161404.
PR	23-JUL-1999;	99US-0145224.	PR	25-OCT-1999;	99US-0161405.
PR	26-JUL-1999;	99US-0145224.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161406.
PR	27-JUL-1999;	99US-0145918.			

PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 100.0%; Score 21; DB 21; Length 39;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTSG 4
Db 20 dtsg 23

RESULT 15

AG57664
ID AG57664 standard; Protein: 54 AA.

XX AC AAG57664;

XX DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 74337.

XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 200EP-0301439.

XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135622.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139889.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.

PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 07-AUG-1999; 99US-0147303.
PR 08-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 14-AUG-1999; 99US-0148565.
PR 15-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149226.
PR 19-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 21-AUG-1999; 99US-0149929.
PR 22-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 24-AUG-1999; 99US-0150566.
PR 25-AUG-1999; 99US-0150884.
PR 26-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 28-AUG-1999; 99US-0151080.
PR 29-AUG-1999; 99US-0151303.
PR 30-AUG-1999; 99US-0151338.
PR 31-AUG-1999; 99US-0151930.
PR 01-SEP-1999; 99US-0152363.
PR 02-SEP-1999; 99US-0153070.
PR 03-SEP-1999; 99US-0153758.
PR 04-SEP-1999; 99US-0154018.
PR 05-SEP-1999; 99US-0154039.
PR 06-SEP-1999; 99US-0154779.
PR 07-SEP-1999; 99US-0155139.
PR 08-SEP-1999; 99US-0155486.
PR 09-SEP-1999; 99US-0155659.
PR 10-SEP-1999; 99US-0156458.
PR 11-SEP-1999; 99US-0156596.
PR 12-SEP-1999; 99US-0157117.
PR 13-SEP-1999; 99US-0157753.
PR 14-SEP-1999; 99US-0157865.
PR 15-SEP-1999; 99US-0158029.
PR 16-SEP-1999; 99US-0158232.
PR 17-SEP-1999; 99US-0158369.
PR 18-SEP-1999; 99US-0159293.
PR 19-SEP-1999; 99US-0159294.
PR 20-SEP-1999; 99US-0159295.
PR 21-SEP-1999; 99US-0159329.
PR 22-SEP-1999; 99US-0159330.
PR 23-SEP-1999; 99US-0159331.
PR 24-SEP-1999; 99US-0159637.
PR 25-SEP-1999; 99US-0159638.
PR 26-SEP-1999; 99US-0159584.
PR 27-SEP-1999; 99US-0160741.
PR 28-SEP-1999; 99US-0160767.
PR 29-SEP-1999; 99US-0160788.
PR 01-OCT-1999; 99US-0160770.
PR 02-OCT-1999; 99US-0160814.
PR 03-OCT-1999; 99US-0160815.
PR 04-OCT-1999; 99US-0160980.
PR 05-OCT-1999; 99US-0160981.
PR 06-OCT-1999; 99US-0160989.
PR 07-OCT-1999; 99US-0161404.
PR 08-OCT-1999; 99US-0161405.
PR 09-OCT-1999; 99US-0161406.
PR 10-OCT-1999; 99US-0161359.
PR 11-OCT-1999; 99US-0161360.
PR 12-OCT-1999; 99US-0161361.
PR 13-OCT-1999; 99US-0161920.
PR 14-OCT-1999; 99US-0161992.
PR 15-OCT-1999; 99US-0161993.
PR 16-OCT-1999; 99US-0162142.

Query Match
Best Local Similarity 100.0%; Score 21; DB 21; Length 54;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTSG 4
Db 21 dtsg 24

Search completed: September 6, 2001, 16:43:29
Job time: 358 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:49:48 ; Search time 231.42 Seconds
(without alignments)
6.861 Million cell updates/sec

Title: US-09-603-713-7
Perfect score: 55
Sequence: 1 KGVVIAIVK 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL16:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_unclassified:*
- 13: sp_vertebrate:*
- 14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	ID	Description
1	45	81.8	49	6	O97917	O97917 bos taurus
2	45	81.8	79	11	O35463	O35463 cricetus
3	45	81.8	82	4	P78438	P78438 homo sapien
4	45	81.8	97	4	Q13778	Q13778 homo sapien
5	45	81.8	534	13	O93296	O93296 gallus gall
6	45	81.8	695	11	P97487	P97487 mus musculus
7	45	81.8	695	11	Q60496	Q60496 cavia porce
8	45	81.8	695	13	O9DGJ8	O9DGJ8 gallus gall
9	45	81.8	699	13	O57394	O57394 narke japon
10	45	81.8	737	13	O93279	O93279 fuqu rubrip
11	45	81.8	747	13	O91963	O91963 xenopus lae
12	45	81.8	751	13	Q9DG37	Q9DG37 gallus gall
13	45	81.8	770	6	Q9TU10	Q9TU10 sus scrofa
14	45	81.8	780	13	O73683	O73683 tetraodon f
15	44	80.0	82	4	Q16019	Q16019 homo sapien
16	44	80.0	612	13	O919E7	O919E7 brachydanio
17	41	74.5	20	4	O9UCB6	O9UCB6 homo sapien
18	41	74.5	82	4	Q16014	Q16014 homo sapien
19	41	74.5	82	4	Q16020	Q16020 homo sapien

20	40	72.7	311	2	O9KWM8	O9Kwm8 pseudomonas
21	40	72.7	569	13	O9PVL1	O9pvl1 gallus gall
22	39	70.9	438	2	O9PRC7	O9prc7 ureaplasma
23	38	69.1	155	1	O9HRS6	O9hrs6 halobacteri
24	38	69.1	430	2	O50505	O50505 streptomyce
25	37	67.3	147	1	O26829	O26829 methanobact
26	37	67.3	256	2	O9X9D0	O9x9d0 salmoneilla
27	37	67.3	395	2	O30627	O30627 escherichia
28	37	67.3	700	4	O9P244	O9p244 homo sapien
29	37	67.3	832	4	O9ULH4	O9ulh4 homo sapien
30	37	67.3	1196	10	O9MOY7	O9moy7 arabidopsis
31	36	65.5	164	10	O9FRP5	O9frf5 oryza sativ
32	36	65.5	242	2	Q47003	Q47003 escherichia
33	36	65.5	242	2	Q46999	Q46999 escherichia
34	36	65.5	265	1	O9HLG4	O9hl94 thermoplasma
35	36	65.5	427	2	O9KSI1	O9ksil vibrio chol
36	36	65.5	479	2	O9F622	O9f622 helicobacte
37	35	63.6	217	10	O9LPU1	O9lpul arabidopsis
38	35	63.6	236	1	O27500	O27500 methanobact
39	35	63.6	298	2	O9PMCS	O9pmcs campylobact
40	35	63.6	300	10	O9SI03	O9si03 arabidopsis
41	35	63.6	339	2	O69943	O69943 streptomyce
42	35	63.6	372	8	O9T7K9	O9t7k9 crassostrea
43	35	63.6	451	4	O9P2G9	O9p2g9 homo sapien
44	35	63.6	510	5	Q9U641	Q9u641 dictyosteli
45	35	63.6	740	3	Q03455	Q03455 saccharomyc

ALIGNMENTS

RESULT 1

O97917 ID O97917 PRELIMINARY: PRT: 49 AA.
AC O97917;
DT 01-MAY-1995 (TREMBLrel. 10, Created)
DT 01-MAY-1995 (TREMBLrel. 10, Last sequence update)
DT 01-MAY-2001 (TREMBLrel. 16, Last annotation update)
DE AMYLOID PRECURSOR PROTEIN (FRAGMENT).
GN APP.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Konfortov B.A., Licence V.E., Miller J.R.;
RT "Re-sequencing of DNA from a diverse panel of cattle reveals frequent polymorphisms in both intron and exon."
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ133033; CAB38017.1;
DR HSSP; P05067; 1BA4.
FT NON_TER 1
FT NON_TER 49
SQ SEQUENCE 49 AA: 5183 MW: 6287463F0559BDED CRC64;

Query Match 81.8%; Score 45; DB 6; Length 49;
Best Local Similarity 100.0%; Pred. No. 0.67;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGVIATVIV 11
Db 20 GGVIATVIV 29
|||||

RESULT 2

O35463 ID O35463 PRELIMINARY: PRT: 79 AA.
AC O35463;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)

01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE ALZHEIMER'S AMYLOID BETA PROTEIN PRECURSOR (FRAGMENT).
 GN BETA APP.
 OS Cricetus griseus (Chinese hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Cricetulus.
 OX NCBI_TaxID=10029;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Sambamurti K., Finnix I., Gandhi S.;
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF030413; AAB86608.1; -
 DR HSSP; P05067; 1QCM.
 FT NON_TER 1
 FT NON_TER 79
 SQ SEQUENCE 79 AA; 8538 MW; 37F2C6C3BFF3F597 CRC64;
 Query Match 81.8%; Score 45; DB 11; Length 79;
 Best Local Similarity 100.0%; Pred. No. 1.1;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 GGVVIATVIV 11
 Db 57 GGVVIATVIV 66
 RESULT 3
 ID P78438 PRELIMINARY; PRT; 82 AA.
 AC P78438;
 DT 01-MAY-1997 (TREMBlrel. 03, Created)
 DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE AMYLOID PROTEIN (BETA-AMYLOID PROTEIN) (FRAGMENT).
 GN APP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Medline-89392030; PubMed-2675837;
 RX Johnstone E.M., Chaney M.O., Moore R.E., Ward K.E., Norris F.H.,
 RA Little S.P.;
 RT "Alzheimer's disease amyloid peptide is encoded by two exons and shows
 RT similarity to soybean trypsin inhibitor.";
 RL Biochem. Biophys. Res. Commun. 163:1248-1255(1989).
 RN [2]
 RP SEQUENCE OF 19-48 FROM N.A.
 RX Medline-87120329; PubMed-2949367;
 RA Tanzi R.E., Gusella J.F., Watkins P.C., Bruns G.A., George-Hyslop P.,
 RA Van Keuren M.L., Patterson D., Pagan S., Kurnit D.M., Neve R.L.;
 RT "Amyloid beta protein gene: cDNA, mRNA distribution, and genetic
 RT linkage near the Alzheimer locus.";
 RL Science 235:880-884(1987).
 RN [3]
 RP SEQUENCE OF 32-63 FROM N.A.
 RX Medline-93035397; PubMed-1415269;
 RA Kamino K., Orr H.T., Payami H., Wijsman E.M., Alonso M.E., Pulst S.M.,
 RA Anderson L., O'dahl S., Nemens E., White J.A.;
 RT "Linkage and mutational analysis of familial Alzheimer disease
 RT kindreds for the APP gene region.";
 RL Am. J. Hum. Genet. 51:998-1014(1992).
 DR EMBL; M28270; AAA51768.1; -
 DR EMBL; M29269; AAA51768.1; JOINED.
 DR EMBL; M15532; AAA51564.1; -
 DR EMBL; S45136; AAB23646.1; -
 DR HSSP; P05067; 1BA4.
 FT NON_TER 1
 FT NON_TER 82 AA; 8994 MW; 8DA9E42B813A070E CRC64;
 SQ SEQUENCE 82 AA; 8994 MW; 8DA9E42B813A070E CRC64;

Query Match 81.8%; Score 45; DB 4; Length 82;
 Best Local Similarity 100.0%; Pred. No. 1.1;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 GGVVIATVIV 11
 Db 53 GGVVIATVIV 62
 RESULT 4
 ID Q13778 PRELIMINARY; PRT; 97 AA.
 AC Q13778;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE AMYLOID PROTEIN (AD-AP) (FRAGMENT).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Medline-87120328; PubMed-3810169;
 RX Goldgaber D., Lerman M.I., McBride O.W., Saffioti U., Gajdusek D.C.;
 RT "Characterization and chromosomal localization of a cDNA encoding
 RT brain amyloid of Alzheimer's disease.";
 RL Science 235:877-880(1987).
 DR EMBL; M15533; AAA35540.1; -
 DR HSSP; P05067; 1BA4.
 DR InterPro; IPR001868; -
 DR PRINTS; PR00203; AMYLOIDA4.
 DR NON_TER 1
 SQ SEQUENCE 97 AA; 10884 MW; E528CDB448DE474E CRC64;
 Query Match 81.8%; Score 45; DB 4; Length 97;
 Best Local Similarity 100.0%; Pred. No. 1.4;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 GGVVIATVIV 11
 Db 35 GGVVIATVIV 44
 RESULT 5
 ID O93296 PRELIMINARY; PRT; 534 AA.
 AC O93296;
 DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE AMYLOID PRECURSOR PROTEIN (FRAGMENT).
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Barnes N.Y., Ling L., Yoshikawa K., Schwartz L.M., Oppenheim R.W.,
 RA Milligan C.E.;
 RT "Increased production of amyloid precursor protein provides a
 RT substrate for Caspase 3 in dying motoneurons.";
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF042098; AAC25052.1; -
 DR HSSP; P05067; 1BA4.
 DR InterPro; IPR001868; -
 DR PRINTS; PR00203; AMYLOIDA4.
 DR PROSITE; PS00319; A4_EXTRA; 1.
 DR PROSITE; PS00320; A4_INTRA; 1.
 FT NON_TER 1

SQ SEQUENCE 534 AA; 60597 MW; FB53ECC2E66D4C92 CRC64;
 Query Match 81.8%; Score 45; DB 13; Length 534;
 Best Local Similarity 100.0%; Pred. No. 8;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 2 GGVIATVIV 11
 Db 472 GGVIATVIV 481
 RESULT 6
 ID P97487 PRELIMINARY; PRT; 695 AA.
 AC P97487; P97487;
 DT 01-MAY-1997 (Tremblrel. 03, Created)
 DT 01-MAR-1997 (Tremblrel. 03, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE HIPPOCAMPAL AMYLOID PROTEIN.
 GN APP.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SAMP8; TISSUE=HIPPOCAMPUS;
 RA Flood J.F., Kumar V.B., Sasser T., Word I., Morley J.E.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 581-662 FROM N.A.
 RC STRAIN=129SV;
 RA Wragg M.A., Busfield F., Duff K., Korenblat K., Capecci M.,
 RA Loring J.F., Goate A.M.;
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U84012; A841502.1; -;
 DR EMBL; U82624; A840919.1; -;
 DR HSSP; P05067; 10CM.
 DR InterPro; IPR001868; -;
 DR Pfam; PF02177; A4_EXTRA; 1.
 DR PRINTS; PR00203; AMYLOIDA4.
 DR PROSITE; PS00319; A4_EXTRA; 1.
 DR PROSITE; PS00320; A4_INTRA; 1.
 DR SMART; SM00006; A4_EXTRA; 1.
 SQ SEQUENCE 695 AA; 78414 MW; 9A5FBE2ED261236E CRC64;
 Query Match 81.8%; Score 45; DB 11; Length 695;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 2 GGVIATVIV 11
 Db 633 GGVIATVIV 642
 RESULT 7
 ID Q60496 PRELIMINARY; PRT; 695 AA.
 AC Q60496;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN.
 OS Cavia porcellus (Guinea pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
 OX NCBI_TaxID=10141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RA Beck M., Mueller D., Bigl V.;

RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: FUNCTIONAL NEURONAL RECEPTOR WHICH COUPLES TO
 CC INTRACELLULAR SIGNALING PATHWAY THROUGH THE GTP-BINDING PROTEIN
 CC G(O).
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -!- SIMILARITY: TO OTHER SPECIES APP ANALOGUES.
 DR EMBL; X97631; CAA66230.1; -;
 DR HSSP; P05067; 1BA4.
 DR InterPro; IPR001868; -;
 DR Pfam; PF02177; A4_EXTRA; 1.
 DR PRINTS; PR00203; AMYLOIDA4.
 DR PROSITE; PS00319; A4_EXTRA; 1.
 DR PROSITE; PS00320; A4_INTRA; 1.
 DR SMART; SM00006; A4_EXTRA; 1.
 SQ SEQUENCE 695 AA; 78701 MW; 5196A0C4017F16AB CRC64;
 Query Match 81.8%; Score 45; DB 11; Length 695;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 2 GGVIATVIV 11
 Db 633 GGVIATVIV 642
 RESULT 8
 ID Q9DGJ8 PRELIMINARY; PRT; 695 AA.
 AC Q9DGJ8;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE BETA-AMYLOID PRECURSOR PROTEIN 695 ISOFORM.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Sarasa M., Rodoloso A., Sorribas V.;
 RT "Cloning of full-length chicken beta-amyloid precursor protein
 RT isoforms.";
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF289218; AAG00593.1; -;
 SQ SEQUENCE 695 AA; 78565 MW; F201ED02AEC86D95 CRC64;
 Query Match 81.8%; Score 45; DB 13; Length 695;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 2 GGVIATVIV 11
 Db 633 GGVIATVIV 642
 RESULT 9
 ID O57394 PRELIMINARY; PRT; 699 AA.
 AC O57394;
 DT 01-JUN-1998 (Tremblrel. 06, Created)
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE EL AMYLOID PRECURSOR PROTEIN 699.
 GN EL APP699.
 OS Narke japonica (Electric ray).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
 OC Elasmobranchii; Squala; Hypnosqualea; Pristiogorae; Batoidae;
 OC Torpediniformes; Narcinoidae; Naridae; Narke.
 OX NCBI_TaxID=62965;
 RN [1]

us-09-603-713-7.rspt

Fri Sep 7 10:58:49 2001

```

RP SEQUENCE FROM N.A.
RC TISSUE-ELECTRIC LOBE;
RA Iijima K., Lee D., Okutsu J., Tomita S., Hirashima N., Kirino Y.,
RA Suzuki T.;
RL Biochem. J. 0:0-0(1998).
DR EMBL; AB005544; BAA24230.1; -.
DR HSSP; P05067; 1QCM.
DR InterPro; IPR001868; -.
DR Pfam; PF02177; A4_EXTRA; 1.
DR PRINTS; PR00203; AMYLOIDA4.
DR PROSITE; PS00319; A4_EXTRA; 1.
DR PROSITE; PS00320; A4_INTRA; 1.
DR SMART; SM00006; A4_EXTRA; 1.
SQ SEQUENCE 699 AA; 78879 MW; 952915C309D50E5C CRC64;

Query Match 81.8%; Score 45; DB 13; Length 699;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGVIATVIV 11
Db 637 GGVIATVIV 646

RESULT 10
O93279 PRELIMINARY; PRT; 737 AA.
AC O93279;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE BETA-AMYLOID PRECURSOR PROTEIN.
GN App. rubripes (Japanese pufferfish) (Takifugu rubripes).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98252138; PubMed=9595080;
RA Villard L., Tassone F., Cnognorac-Jurcevic T., Clancy K., Gardiner K.;
RT "Analysis of pufferfish homologues of the AT-rich human App gene.";
RL Gene 210:17-24(1998).
CC -1- SIMILARITY: TO PANCREATIC TRYPsin INHIBITOR (KUNITZ) DOMAIN.
DR EMBL; AF090120; AAD13392.1; -.
DR HSSP; P05067; 1QCM.
DR InterPro; IPR001868; -.
DR InterPro; IPR002223; -.
DR Pfam; PF00014; Kunitz_BPTI; 1.
DR Pfam; PF02177; A4_EXTRA; 1.
DR PRINTS; PR00203; AMYLOIDA4.
DR PRINTS; PR00759; BASICPTASE.
DR PROSITE; PS00320; A4_INTRA; 1.
DR PROSITE; PS00280; BPTI_KUNITZ; 1.
DR SMART; SM00131; KU; 1.
DR Serine protease inhibitor.
KW SEQUENCE 737 AA; 82856 MW; 6FAD01E2E3B2B7E2 CRC64;

Query Match 81.8%; Score 45; DB 13; Length 737;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGVIATVIV 11
Db 675 GGVIATVIV 684

RESULT 11-
Q91963

```

```

ID Q91963 PRELIMINARY; PRT; 747 AA.
AC Q91963;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE APP747. laevis (African clawed frog).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93129227; PubMed=1282805;
RA Okado H., Okamoto H.;
RT "A Xenopus homologue of the human beta-amyloid precursor protein: development regulation of its gene expression.";
RL Biochem. Biophys. Res. Commun. 189:1561-1568(1992).
CC -1- SIMILARITY: TO PANCREATIC TRYPsin INHIBITOR (KUNITZ) DOMAIN.
DR EMBL; S52417; AAB24853.1; -.
DR HSSP; P05067; 1QCM.
DR InterPro; IPR001868; -.
DR InterPro; IPR002223; -.
DR Pfam; PF00014; Kunitz_BPTI; 1.
DR Pfam; PF02177; A4_EXTRA; 1.
DR PRINTS; PR00203; AMYLOIDA4.
DR PRINTS; PR00759; BASICPTASE.
DR PROSITE; PS00319; A4_EXTRA; 1.
DR PROSITE; PS00320; A4_INTRA; 1.
DR PROSITE; PS00280; BPTI_KUNITZ; 1.
DR SMART; SM00131; KU; 1.
DR Serine protease inhibitor.
KW SEQUENCE 747 AA; 84892 MW; A75E81885681D948 CRC64;

Query Match 81.8%; Score 45; DB 13; Length 747;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGVIATVIV 11
Db 685 GGVIATVIV 694

RESULT 12
Q9DGJ7 PRELIMINARY; PRT; 751 AA.
AC Q9DGJ7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE BETA-AMYLOID PRECURSOR PROTEIN 751 ISOFORM.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Sakasa M., Rodolosse A., Sorribas V.;
RT "Cloning of full-length chicken beta-amyloid precursor protein isoforms.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF289219; AAG00594.1; -.
DR EMBL; AF289219; AAG00594.1; -.
SQ SEQUENCE 751 AA; 84705 MW; E78E9413A8033D84 CRC64;

Query Match 81.8%; Score 45; DB 13; Length 751;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGVIATVIV 11
Db 685 GGVIATVIV 694

```


Db 689 GGVVIATVIV 698

RESULT 13

Q9TUI0 PRELIMINARY; PRT; 770 AA.
ID Q9TUI0; AC Q9TUI0; DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE AMYLOID PRECURSOR PROTEIN.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA Kimura A., Takahashi T.;
RT "Amyloid Precursor Protein 770.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO PANCREATIC TRYPsin INHIBITOR (KUNITZ) DOMAIN.
DR EMBL; AB032550; BAA84580.1; -.
DR HSSP; P05067; 1AAP.
DR InterPro; IPR001868; -.
DR InterPro; IPR002223; -.
DR Pfam; PF00014; Kunitz_BPTI; 1.
DR Pfam; PF02177; A4_EXTRA; 1.
DR PRINTS; PR00203; AMYLOIDA4.
DR PRINTS; PR00759; BASICPTASE.
DR PROSITE; PS00319; A4_EXTRA; 1.
DR PROSITE; PS00320; A4_INTRA; 1.
DR PROSITE; PS00280; BPTI_KUNITZ; 1.
DR SMART; SM00006; A4_EXTRA; 1.
KW Serine protease inhibitor.
SQ SEQUENCE 770 AA; 86961 MW; 5F7A1DCB2BCC583E CRC64;

Query Match 81.8%; Score 45; DB 6; Length 770;

Best Local Similarity 100.0%; Pred. No. 12;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGVVIATVIV 11

Db 708 GGVVIATVIV 717

RESULT 14

O73683 PRELIMINARY; PRT; 780 AA.
ID O73683; AC O73683; DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE AMYLOID PRECURSOR PROTEIN.
OS Tetraodon fluviatilis (Puffer fish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodon.
OX NCBI_TaxID=47145;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-98252138; PubMed-9599080;
RX Villard L., Tassone F., Crnogorac-Jurcevic T., Clancy K., Gardiner K.;
RT "Analysis of pufferfish homologues of the AT-rich human APP gene.";
RL Gene 210:17-24(1998).
CC -1- SIMILARITY: TO PANCREATIC TRYPsin INHIBITOR (KUNITZ) DOMAIN.
DR EMBL; AF018165; AAC41275.1; -.
DR HSSP; P05067; 1QCM.
DR InterPro; IPR001868; -.
DR InterPro; IPR002223; -.
DR Pfam; PF00014; Kunitz_BPTI; 1.
DR Pfam; PF02177; A4_EXTRA; 1.

DR PRINTS; PR00203; AMYLOIDA4.
DR PRINTS; PR00759; BASICPTASE.
DR PROSITE; PS00319; A4_EXTRA; 1.
DR PROSITE; PS00320; A4_INTRA; 1.
DR SMART; SM00131; KU; 1.
KW Serine protease inhibitor.
SQ SEQUENCE 780 AA; 88238 MW; 60071BE94520191D CRC64;

Query Match 81.8%; Score 45; DB 13; Length 780;

Best Local Similarity 100.0%; Pred. No. 12;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGVVIATVIV 11

Db 718 GGVVIATVIV 727

RESULT 15

Q16019 PRELIMINARY; PRT; 82 AA.
ID Q16019; AC Q16019; DT 01-NOV-1995 (TREMBLrel. 01, Created)
DT 01-NOV-1995 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE BETA-AMYLOID PEPTIDE PRECURSOR (FRAGMENT).
GN BETA APP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-93236601; PubMed-8476439;
RA Denman R.B., Rosenczwaig R., Miller D.L.;
RT "A system for studying the effect(s) of familial Alzheimer disease mutations on the processing of the beta-amyloid peptide precursor.";
RL Biochem. Biophys. Res. Commun. 192:96-103(1993).
DR EMBL; S61380; AAB26264.2; -.
DR HSSP; P05067; 1BA4.
DR NON_TER 1
FT NON_TER 82
SQ SEQUENCE 82 AA; 8938 MW; F534AA50E579230A CRC64;

Query Match 80.0%; Score 44; DB 4; Length 82;

Best Local Similarity 90.0%; Pred. No. 1.7;

Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGVVIATVIV 11

Db 54 GGVVIATVIV 63

Search completed: September 6, 2001, 16:49:48

Job time: 732 sec

us-09-603-713-7.rspt

Fri Sep 7 10:58:49 2001

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	45	81.8	57	1	A4_PIG	Q29023 ursus scrofa
2	45	81.8	51	1	A4_URSMA	Q29149 ursus marit
3	45	81.8	58	1	A4_CANFA	Q28280 canis famil
4	45	81.8	58	1	A4_RABIT	Q28748 oryctolagus
5	45	81.8	58	1	A4_SHEEP	Q28757 ovis aries
6	45	81.8	59	1	A4_BOVIN	Q28053 bos taurus
7	45	81.8	751	1	A4_SAISC	Q95241 salmirl sci
8	45	81.8	770	1	A4_HUMAN	P05067 homo sapien
9	45	81.8	770	1	A4_MOUSE	P12023 mus musculu
10	45	81.8	770	1	A4_RAT	P08592 rattus norv
11	36	65.5	242	1	TRKL_ECOLI	P41066 escherichia
12	36	65.5	536	1	FLIF_BACSU	P23447 bacillus su
13	36	65.5	538	1	ROGO_MOUSE	P08848 mus musculu
14	36	65.5	657	1	DCTS_RHOCA	P37739 rhodobacter
15	36	65.5	727	1	PECL_MOUSE	Q08481 mus musculu
16	35	63.6	729	1	VGNB_APMV	Q02941 andean pota
17	34	61.8	231	1	NUAM_BOTBI	Q03698 bothriopsis
18	34	61.8	231	1	NUAM_BOTER	Q03699 bothrops er
19	34	61.8	231	1	NUAM_LACMU	P92649 lachesis mu
20	34	61.8	231	1	NUAM_PORHY	Q03763 porthidium
21	34	61.8	375	1	MRKL_YEAST	P50873 saccharomyc
22	34	61.8	465	1	CGAL_HUMAN	P78396 homo saplen
23	34	61.8	1004	1	ATNA_ARTSF	P28774 artemia san
24	34	61.8	2035	1	HFCL_HUMAN	P51610 homo saplen
25	34	61.8	3412	1	POLG_TBESV	P07720 t genome pos
26	33	60.0	65	1	VJET_ECOLI	P39289 escherichia
27	33	60.0	110	1	YNF7_YEAST	P53948 saccharomyc
28	33	60.0	122	1	RL14_HELPJ	Q923s2 helicobacte
29	33	60.0	183	1	PYRE_THETH	Q60016 thermus aqu
30	33	60.0	220	1	Y085_MYCTU	Q10882 mycobacteri
31	33	60.0	244	1	YRT2_CAEEL	Q10045 caenorhabdi
32	33	60.0	275	1	NUIM_MYTEE	Q00860 mytilus edu
33	33	60.0	295	1	PP1B_MYCLE	P46697 mycobacteri


```
RL Brain Res. Mol. Brain Res. 10:299-305(1991).
CC -!- FUNCTION: FUNCTIONAL NEURONAL RECEPTOR WHICH COUPLES TO
CC INTRACELLULAR SIGNALING PATHWAY THROUGH THE GTP-BINDING PROTEIN
CC G(O) (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE APP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X56129; CAA39594.1; -.
DR HSSP; P05067; IAML.
DR InterPro; IPR001868; -.
DR PROSITE; PS00319; A4_EXTRA; PARTIAL.
DR PROSITE; PS00320; A4_INTRA; PARTIAL.
KW Glycoprotein; Amyloid; Neurone; Transmembrane.
FT NON_TER 1 1
FT CHAIN 6 48 BETA-AMYLOID PROTEIN (POTENTIAL).
FT DOMAIN <1 33 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 34 57 POTENTIAL.
FT DOMAIN 58 >58 CYTOPLASMIC (POTENTIAL).
FT NON_TER 58 58
FT SEQUENCE 58 AA; 6300 MW; F434209D8EBA82D CRC64;
DR EMBL; X56129; CAA39594.1; -.
DR HSSP; P05067; IAML.
DR InterPro; IPR001868; -.
DR PROSITE; PS00319; A4_EXTRA; PARTIAL.
DR PROSITE; PS00320; A4_INTRA; PARTIAL.
KW Glycoprotein; Amyloid; Neurone; Transmembrane.
FT NON_TER 1 1
FT CHAIN 6 48 BETA-AMYLOID PROTEIN (POTENTIAL).
FT DOMAIN <1 33 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 34 57 POTENTIAL.
FT DOMAIN 58 >58 CYTOPLASMIC (POTENTIAL).
FT NON_TER 58 58
FT SEQUENCE 58 AA; 6300 MW; F434209D8EBA82D CRC64;

Query Match 81.8%; Score 45; DB 1; Length 58;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGVVIATVIV 11
DB 42 GGVVIATVIV 51

RESULT 5
A4_SHEEP
ID A4_SHEEP STANDARD; PRT; 58 AA.
AC Q28757;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN HOMOLOG [CONTAINS: BETA-AMYLOID
DE PROTEIN (BETA-APP) (A-BETA)] (FRAGMENT).
GN APP.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92017079; PubMed=1656157;
RA Johnstone E.M., Chaney M.O., Norris F.H., Pascual R., Little S.P.;
RT "Conservation of the sequence of the Alzheimer's disease amyloid
RT peptide in dog, polar bear and five other mammals by cross-species
RT polymerase chain reaction analysis.";
RL Brain Res. Mol. Brain Res. 10:299-305(1991).
CC -!- FUNCTION: FUNCTIONAL NEURONAL RECEPTOR WHICH COUPLES TO
CC INTRACELLULAR SIGNALING PATHWAY THROUGH THE GTP-BINDING PROTEIN
CC G(O) (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE APP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X56130; CAA39595.1; -.
DR HSSP; P05067; IAML.
DR InterPro; IPR001868; -.
DR PROSITE; PS00319; A4_EXTRA; PARTIAL.
DR PROSITE; PS00320; A4_INTRA; PARTIAL.
KW Glycoprotein; Amyloid; Neurone; Transmembrane.
FT NON_TER 1 1
FT CHAIN 6 48 BETA-AMYLOID PROTEIN (POTENTIAL).
FT DOMAIN <1 33 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 34 57 POTENTIAL.
FT DOMAIN 58 >58 CYTOPLASMIC (POTENTIAL).
FT NON_TER 58 58
FT SEQUENCE 58 AA; 6300 MW; F434209D8EBA82D CRC64;

Query Match 81.8%; Score 45; DB 1; Length 58;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGVVIATVIV 11
DB 42 GGVVIATVIV 51

RESULT 6
A4_BOVIN
ID A4_BOVIN STANDARD; PRT; 59 AA.
AC Q28053;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN HOMOLOG [CONTAINS: BETA-AMYLOID
DE PROTEIN (BETA-APP) (A-BETA)] (FRAGMENT).
GN APP.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92017079; PubMed=1656157;
RA Johnstone E.M., Chaney M.O., Norris F.H., Pascual R., Little S.P.;
RT "Conservation of the sequence of the Alzheimer's disease amyloid
RT peptide in dog, polar bear and five other mammals by cross-species
RT polymerase chain reaction analysis.";
RL Brain Res. Mol. Brain Res. 10:299-305(1991).
CC -!- FUNCTION: FUNCTIONAL NEURONAL RECEPTOR WHICH COUPLES TO
CC INTRACELLULAR SIGNALING PATHWAY THROUGH THE GTP-BINDING PROTEIN
CC G(O) (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE APP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X56124; CAA39589.1; -.
DR EMBL; X56126; CAA39591.1; -.
DR HSSP; P05067; IAML.
DR InterPro; IPR001868; -.
DR PROSITE; PS00319; A4_EXTRA; PARTIAL.
DR PROSITE; PS00320; A4_INTRA; PARTIAL.
KW Glycoprotein; Amyloid; Neurone; Transmembrane.
```

```

FT NON_TER 1 1 BETA-AMYLOID PROTEIN (POTENTIAL).
FT CHAIN 7 49 EXTRACELLULAR (POTENTIAL).
FT DOMAIN <1 34 POTENTIAL.
FT TRANSMEM 35 58 CYTOPLASMIC (POTENTIAL).
FT NON_TER 59 >59 F43469D488A2E12D CRC64;
SQ SEQUENCE 59 AA; 6414 MW; 81.8%; Score 45; DB 1; Length 59;

Query Match
Best Local Similarity 100.0%; Pred. No. 0.2;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGVVIATVIV 11
DB 43 GGVVIATVIV 52

RESULT 7
A4_SAISC STANDARD; PRT; 751 AA.
AC Q95241;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN PRECURSOR [CONTAINS: BETA-
DE AMYLOID PROTEIN (BETA-APP) (A-BETA)].
GN APP.
OS Saimiri sciureus (Common squirrel monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae; Saimiri.
OX NCBI_TaxID=9521;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver, and Kidney;
RX MEDLINE=96108492; PubMed=8532114;
RA Levy E., Amorim A., Frangione B., Walker L.C.;
RT "Beta-amyloid precursor protein gene in squirrel monkeys with
RT cerebral amyloid angiopathy.";
RL Neurobiol. Aging 16:805-808(1995).
CC -!- FUNCTION: FUNCTIONAL NEURONAL RECEPTOR WHICH COUPLES TO
CC G(I).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- DOMAIN: THE CLATHRIN-BINDING SITE IS ESSENTIAL FOR ITS ASSOCIATION
CC WITH X11-ALPHA, -BETA, AND -GAMMA. THE SEQUENCE SPECIFIC
CC RECOGNITION EXTENDS TO PEPTIDE RESIDUES THAT ARE C-TERMINAL TO THE
CC NPXY MOTIF. THIS INTERACTION APPEARS TO BE INDEPENDENT OF
CC PHOSPHORYLATION (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE APP FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; S81024; AAD14347.1; -
CC InterPro; IPR001255; -
CC InterPro; IPR001868; -
CC Pfam; PF00014; Kunitz_BPTI; 1.
CC PRINTS; PR00203; AMYLOIDA4.
CC PRINTS; PR00204; BETAAMYLOID.
CC PRINTS; PR00759; BASICPTASE.
CC PROSITE; PS00319; A4_EXTRA; 1.
CC PROSITE; PS00320; A4_INTRA; 1.
CC PROSITE; PS00280; BPTI_KUNITZ_1; 1.
CC PROSITE; PS00279; BPTI_KUNITZ_2; 1.
CC GlycoProtein; Amyloid; Neurone; Transmembrane; Alternative splicing;
CC Signal; Serine protease inhibitor.

```

```

FT SIGNAL 1 17 BY SIMILARITY.
FT CHAIN 18 751 A4 PROTEIN.
FT CHAIN 653 695 BETA-AMYLOID PROTEIN (POTENTIAL).
FT DOMAIN 18 680 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 681 704 POTENTIAL.
FT DOMAIN 705 751 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 287 345 BPTI/KUNITZ INHIBITOR.
FT SITE 740 743 CLATHRIN-BINDING (BY SIMILARITY).
FT ACT_SITE 301 302 REACTIVE BOND.
FT DISULFID 291 341 BY SIMILARITY.
FT DISULFID 300 324 BY SIMILARITY.
FT CARBOHYD 316 337 BY SIMILARITY.
FT CARBOHYD 523 523 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 552 552 N-LINKED (GLCNAC. . .) (PROBABLE).
SQ SEQUENCE 751 AA; 84893 MW; 6C3B431089569049 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGVVIATVIV 11
DB 689 GGVVIATVIV 698

RESULT 8
A4_HUMAN STANDARD; PRT; 770 AA.
AC P05067; P09000; Q16011;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN PRECURSOR (PROTEASE NEXIN-II)
DE (PN-II) (APP) [CONTAINS: BETA-AMYLOID PROTEIN (BETA-APP) (A-BETA)].
GN APP OR A4 OR CVAP OR ADL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=87144572; PubMed=2881207;
RA Kang J., Lemaire H.-G., Unterbeck A., Salbaum J.M., Masters C.L.,
RA Grzeschik K.-H., Multhaup G., Beyreuther K., Mueller-Hill B.;
RT "The precursor of Alzheimer's disease amyloid A4 protein resembles a
RT cell-surface receptor.";
RL Nature 325:733-736(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88122639; PubMed=2893289;
RA Ponte P., Gonzalez-Dewhitt P., Schilling J., Miller J., Hsu D.,
RA Greenberg B., Davis K., Wallace W., Lieberburg I., Fuller F.,
RA Cordell B.;
RT "A new A4 amyloid mRNA contains a domain homologous to serine
RT proteinase inhibitors.";
RL Nature 331:525-527(1988).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=89128427; PubMed=2783775;
RA Lemaire H.G., Salbaum J.M., Multhaup G., Kang J., Bayney R.M.,
RA Unterbeck A., Beyreuther K., Mueller-Hill B.;
RT "The PreA4(695) precursor protein of Alzheimer's disease A4 amyloid
RT is encoded by 16 exons.";
RL Nucleic Acids Res. 17:517-522(1989).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=97263807; PubMed=9108164;
RA Hattori M., Tsukahara F., Furuhashi Y., Tanahashi H., Hirose M.,
RA Saito M., Tsukuni S., Sakaki Y.;
RT "A novel method for making nested deletions and its application for
RT sequencing of a 300 kb region of human APP locus.";

```

RL Nucleic Acids Res. 25:1802-1808(1997).
RN [5]
RP SEQUENCE OF 286-345 AND 365-366 FROM N.A.
RX MEDLINE=88122640; PubMed=2893290;
RA Tanzi R.E., McClatchey A.I., Lamperti E.D., Villa-Komaroff L.,
RA Gusella J.F., Neve R.L.;
RT "Protease inhibitor domain encoded by an amyloid protein precursor
RT mRNA associated with Alzheimer's disease.";
RL Nature 331:528-530(1988).
RN [6]
RP SEQUENCE OF 287-367 FROM N.A.
RX MEDLINE=88122641; PubMed=2893291;
RA Kitaguchi N., Takahashi Y., Tokushima Y., Shiojiri S., Ito H.;
RT "Novel precursor of Alzheimer's disease amyloid protein shows
RT protease inhibitory activity.";
RL Nature 331:530-532(1988).
RN [7]
RP SEQUENCE OF 284-289 AND 365-770 FROM N.A.
RX MEDLINE=87231971; PubMed=3035574;
RA Robakis N.K., Ramakrishna N., Wolfe G., Wisniewski H.M.;
RT "Molecular cloning and characterization of a cDNA encoding the
RT cerebrovascular and the neuritic plaque amyloid peptides.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:4190-4194(1987).
RN [8]
RP SEQUENCE OF 507-770 FROM N.A.
RX MEDLINE=88124954; PubMed=2893379;
RA Zain S.B., Salim M., Chou W.G., Sajdel-Sulkowska E.M., Majocha R.E.,
RA Marotta C.A.;
RT "Molecular cloning of amyloid cDNA derived from mRNA of the Alzheimer
RT disease brain: coding and noncoding regions of the fetal precursor
RT mRNA are expressed in the cortex.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:929-933(1988).
RN [9]
RP SEQUENCE OF 672-681.
RX MEDLINE=88035004; PubMed=3312495;
RA Pardridge W.M., Vinters H.V., Yang J., Eisenberg J., Choi T.B.,
RA Tourtellotte W.W., Huebner V., Shively J.E.;
RT "Amyloid angiopathy of Alzheimer's disease: amino acid composition
RT and partial sequence of a 4,200-dalton peptide isolated from cortical
RT microvessels.";
RL J. Neurochem. 49:1394-1401(1987).
RN [10]
RP SEQUENCE OF 739-770 FROM N.A.
RX MEDLINE=90236318; PubMed=2110105;
RA Yoshikai S.-I., Sasaki H., Doh-Ura K., Furuya H., Sakaki Y.;
RT "Genomic organization of the human amyloid beta-protein precursor
RT gene.";
RL Gene 87:257-263(1990).
RN [11]
RP SEQUENCE OF 1-10 FROM N.A.
RX TISSUE=Liver;
RC MEDLINE=89016647; PubMed=3140222;
RA Schon E.A., Mita S., Sadlock J., Herbert J.;
RT "A cDNA specifying the human amyloid beta precursor protein (ABPP)
RT encodes a 95-kDa polypeptide.";
RL Nucleic Acids Res. 16:9351-9351(1988).
RN [12]
RP SEQUENCE OF 18-50.
RX MEDLINE=87250462; PubMed=3597385;
RA van Nostrand W.E., Cunningham D.D.;
RT "Purification of protease nexin II from human fibroblasts.";
RL J. Biol. Chem. 262:8508-8514(1987).
RN [13]
RP IDENTITY OF APP WITH NEXIN-II.
RX MEDLINE=89384866; PubMed=2506449;
RA Oltersdorf T., Fritz L.C., Schenk D.B., Lieberburg I.,
RA Johnson-Wood K.L., Beattie E.C., Ward P.J., Blacher R.W., Dovey H.F.,
RA Sinha S.;
RT "The secreted form of the Alzheimer's amyloid precursor protein with
RT the Kunitz domain is protease nexin-II.";
RL Nature 341:144-147(1989).
RN [14]
RP PROTEASE-SPECIFICITY OF INHIBITOR DOMAIN.

RX MEDLINE=90211252; PubMed=1969731;
RA Kido H., Fukutomi A., Schilling J., Wang Y., Cordell B., Katunuma N.;
RT "Protease-specificity of Kunitz inhibitor domain of Alzheimer's
RT disease amyloid protein precursor.";
RL Biochem. Biophys. Res. Commun. 167:716-721(1990).
RN [15]
RP COMPLEX WITH G(O).
RX MEDLINE=93189965; PubMed=8446172;
RA Nishimoto I., Okamoto T., Matsuura Y., Takahashi S., Okamoto T.,
RA Murayama Y., Ogata E.;
RT "Alzheimer amyloid protein precursor complexes with brain GTP-binding
RT protein G(O).";
RL Nature 362:75-79(1993).
RN [16]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 28-133.
RX MEDLINE=99315582; PubMed=10201399;
RA Rossjohn J., Cappai R., Feil S.C., Henry A., McKinstry W.J.,
RA Galatis D., Hesse L., Multhaup G., Beyreuther K., Masters C.L.,
RA Parker M.W.;
RT "Crystal structure of the N-terminal, growth factor-like domain of
RT Alzheimer amyloid precursor protein.";
RL Nat. Struct. Biol. 6:327-331(1999).
RN [17]
RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF 287-344.
RX MEDLINE=91104913; PubMed=2125487;
RA Hynes T.R., Randal M., Kennedy L.A., Eigenbrot C., Kossiakof A.A.;
RT "X-ray crystal structure of the protease inhibitor domain of
RT Alzheimer's amyloid beta-protein precursor.";
RL Biochemistry 29:10018-10022(1990).
RN [18]
RP STRUCTURE BY NMR OF 289-344.
RX MEDLINE=92031488; PubMed=1718421;
RA Heald S.L., Tilton R.F. Jr., Hammond L.S., Lee A., Bayney R.M.,
RA Karmark M.H., Ramabhadran T.V., Dreyer R.N., Davis G., Unterbeck A.,
RA Tamburini P.P.;
RT "Sequential NMR resonance assignment and structure determination of
RT the Kunitz-type inhibitor domain of the Alzheimer's beta-amyloid
RT precursor protein.";
RL Biochemistry 30:10467-10478(1991).
RN [19]
RP STRUCTURE BY NMR OF 672-699.
RX MEDLINE=94281210; PubMed=7516706;
RA Talafous J., Marciniowski K.J., Klopman G., Zagorski M.G.;
RT "Solution structure of residues 1-28 of the amyloid beta-peptide.";
RL Biochemistry 33:7788-7796(1994).
RN [20]
RP STRUCTURE BY NMR OF 696-706.
RX MEDLINE=97128622; PubMed=8973180;
RA Kohno T., Kobayashi K., Maeda T., Sato K., Takashima A.;
RT "Three-dimensional structures of the amyloid beta peptide (25-35) in
RT membrane-mimicking environment.";
RL Biochemistry 35:16094-16104(1996).
RN [21]
RP STRUCTURE BY NMR OF 672-711.
RX MEDLINE=98359783; PubMed=9693002;
RA Coles M., Eicknell W., Watson A.A., Craik D.J.;
RT "Solution structure of amyloid beta-peptide(1-40) in a water-micelle
RT environment. Is the membrane-spanning domain where we think it is?";
RL Biochemistry 37:11064-11077(1998).
RN [22]
RP STRUCTURE BY NMR OF 672-699.
RX MEDLINE=20400066; PubMed=10940222;
RA Poulsen S.-A., Watson A.A., Craik D.J.;
RT "Solution structures in aqueous SDS micelles of two amyloid beta
RT peptides of Abeta(1-28) mutated at the alpha-secretase cleavage
RT site.";
RL J. Struct. Biol. 130:142-152(2000).
RN [23]
RP SIGNAL SEQUENCE CLEAVAGE SITE, AND TOPOLOGY.
RX MEDLINE=88296437; PubMed=2900137;
RA Dykx T., Weidemann A., Multhaup G., Salbaum J.M., Lemaire H.-G.,
RA Kang J., Mueller-Hill B., Masters C.L., Beyreuther K.;
RT "Identification, transmembrane orientation and biogenesis of the

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGVVIATVIV 11
1111111111

Db 708 GGVVIATVIV 717

RESULT 10

A4_RAT
ID A4_RAT STANDARD; PRT; 770 AA.
AC P08592;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN HOMOLOG PRECURSOR
DE (AMYLOIDGENIC GLYCOPROTEIN) (AG).
GN APP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE OF 1-289 AND 365-770 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=88312583; PubMed=2900758;
RA Shivers B.D., Hilbich C., Multhaup G., Salbaum J.M., Beyreuther K.,
Seeburg P.H.;
RT "Alzheimer's disease amyloidogenic glycoprotein: expression pattern
in rat brain suggests a role in cell contact.";
RL EMBO J. 7:1365-1370(1988).
RN [2]
RP SEQUENCE OF 289-364 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=89183625; PubMed=2648331;
RA Kang J., Mueller-Hill B.;
RT The sequence of the two extra exons in rat preA4.";
RL Nucleic Acids Res. 17:2130-2130(1989).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: SIX FORMS OF APP ARE FOUND; APP(395),
APP(563), APP(695), APP(751) AND APP(770) (SHOWN HERE); WHICH ARE
PRODUCED BY ALTERNATIVE SPLICING.
CC -1- DOMAIN: THE CLATHRIN-BINDING SITE IS ESSENTIAL FOR ITS ASSOCIATION
WITH X11-ALPHA, -BETA, AND -GAMMA. THE SEQUENCE SPECIFIC
RECOGNITION EXTENDS TO PEPTIDE RESIDUES THAT ARE C-TERMINAL TO THE
NPXY MOTIF. THIS INTERACTION APPEARS TO BE INDEPENDENT OF
PHOSPHORYLATION (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE APP FAMILY.
CC -1- SIMILARITY: APP(751) AND APP(770) CONTAIN A PROTEASE INHIBITOR
DOMAIN BELONGING TO THE BPTI/KUNITZ FAMILY OF INHIBITORS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; X07648; CAA30488.1; -;
EMBL; X14066; CAA32229.1; -;
PIR; S00550; S00550.
DR PIR; S03607; S03607.
DR InterPro; IPR001255; -;
DR InterPro; IPR001868; -;
DR InterPro; IPR002223; -;
DR Pfam; PF00014; Kunitz_BPTI; 1.
DR PRINTS; PR00203; AMYLOIDA4.
DR PRINTS; PR00204; BETAAMYLOID.
DR PRINTS; PR00759; BASICPTASE.
DR PROSITE; PS00319; A4_EXTRA; 1.
DR PROSITE; PS00320; A4_INTRA; 1.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 1.
DR PROSITE; PS00279; BPTI_KUNITZ_2; 1.

KW Glycoprotein; Amyloid; Neurone; Transmembrane; Signal;
KW Alternative splicing; Serine protease inhibitor.
FT SIGNAL 1 17
FT CHAIN 18 770
FT ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN
FT HOMOLOG.
FT DOMAIN 18 699
FT TRANSNEM 700 723
FT DOMAIN 724 770
FT DOMAIN 673 715
FT DOMAIN 287 345
FT SITE 759 762
FT DISULFID 291 341
FT DISULFID 300 324
FT DISULFID 316 337
FT CARBOHYD 542 542
FT CARBOHYD 571 571
FT VARSPLIC 289 289
FT VARSPLIC 290 364
SQ SEQUENCE 770 AA; 86704 MW; C26C9D6BB2D929A7 CRC64;

Query Match 81.8%; Score 45; DB 1; Length 770;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGVVIATVIV 11
1111111111

Db 708 GGVVIATVIV 717

RESULT 11

TRK1_ECOLI
ID TRK1_ECOLI STANDARD; PRT; 242 AA.
AC P41066;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE TRAK PROTEIN.
GN TRAK.
OS Escherichia coli.
OG Plasmid F.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94359430; PubMed=7915817;
RA Frost L.S., Ippen-Ihler K., Skurray R.A.;
RT "Analysis of the sequence and gene products of the transfer region of
the F sex factor.";
RL Microbiol. Rev. 58:162-210(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / CR63;
RA Shimizu H., Saitoh Y., Suda Y., Uehara K., Sampei G., Mizobuchi K.;
RT "Complete nucleotide sequence of the F plasmid: its implications for
organization and diversification of plasmid genomes.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN F PILUS ASSEMBLY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; U01159; AAC44189.1; -;
EMBL; AP001918; BAA97947.1; -;
DR EcoGene; EG40092; trak.
KW Plasmid; Conjugation.
SQ SEQUENCE 242 AA; 25627 MW; F422AC10F48A1264 CRC64;

KW Flagella; Membrane.
SQ SEQUENCE 536 AA; 59300 MW; 21110D4C7CF1927A CRC64;

Query Match 65.5%; Score 36; DB 1; Length 242;
Best Local Similarity 87.5%; Pred. No. 24;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GGVVIATV 9
|||||
Db 77 GGVVATV 84

RESULT 12
FLIF_BACSU STANDARD; PRT; 536 AA.
ID P23447;
AC 01-NOV-1991 (Rel. 20, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE FLAGELLAR M-RING PROTEIN.
GN FLIF.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91285431; PubMed=1905667;
RA Zuberi A.R., Ying C., Bischoff D.S., Ordal G.W.;
RT "Gene-protein relationships in the flagellar hook-basal body complex
of Bacillus subtilis: sequences of the flgB, flgC, flgG, flfE and
flfF genes.";
RT Mol. Biol. Rep. 23:205-210(1996).
RL Gene 101:23-31(1991).
RN [2]
RP SEQUENCE OF 355-536 FROM N.A.
RC STRAIN=168;
RX MEDLINE=91258343; PubMed=1828465;
RA Albertini A.M., Caramori T., Crabb W.D., Scoffone F., Galizzi A.;
RT "The flag locus of Bacillus subtilis is part of a large operon coding
for flagellar structures, motility functions, and an ATPase-like
polypeptide.";
RT J. Bacteriol. 173:3573-3579(1991).
RN [3]
RP FUNCTION: THE M RING MAY BE ACTIVELY INVOLVED IN ENERGY
TRANSDUCTION.
CC -!- SUBUNIT: THE BASAL BODY CONSTITUTES A MAJOR PORTION OF THE
CC FLAGELLAR ORGANELLE AND CONSISTS OF FOUR RINGS (L,P,S, AND M)
CC MOUNTED ON A CENTRAL ROD. THE M RING IS INTEGRAL TO THE INNER
CC MEMBRANE OF THE CELL AND MAY BE CONNECTED TO THE FLAGELLAR ROD
CC VIA THE S RING. THE S (SUPRAMEMBRANE RING) LIES JUST DISTAL TO
CC THE M RING. THE L AND P RINGS LIE IN THE OUTER MEMBRANE AND THE
CC PERIPLASMIC SPACE, RESPECTIVELY.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC, MEMBRANE-ASSOCIATED.
CC -!- SIMILARITY: BELONGS TO THE FLIF FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M54965; -, NOT_ANNOTATED_CDS.
DR EMBL; X56049; CAA39520.1; -.
DR EMBL; Z99112; CAB13494.1; -.
DR PIR; JG0022; JG0022.
DR PIR; S14494; S14494.
DR PIR; A42365; A42365.
DR Subtilist; BG10240; flif.
DR InterPro; IPR000067; -.
DR InterPro; IPR001899; -.
DR InterPro; IPR002920; -.
DR Pfam; PF01514; YscJ_Flif; 1.
DR PRINTS; PR01009; FLGMRINGFLIF.

KW Flagella; Membrane.
SQ SEQUENCE 536 AA; 59300 MW; 21110D4C7CF1927A CRC64;

Query Match 65.5%; Score 36; DB 1; Length 536;
Best Local Similarity 70.0%; Pred. No. 46;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GGVVIATVIV 11
|||||
Db 456 GGVIIAAIIV 465

RESULT 13
RO60_MOUSE STANDARD; PRT; 538 AA.
ID RO60_MOUSE
AC O08848; Q9QVD8;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE 60-KDA SS-A/RO RIBONUCLEOPROTEIN (60 KDA RO PROTEIN) (60 KDA
DE RIBONUCLEOPROTEIN RO) (RORNP).
GN SRA2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=97266462; PubMed=9112230;
RA Wang D., Buyon J.P., Chan E.K.L.;
RT "Cloning and expression of mouse 60 kDa ribonucleoprotein SS-A/Ro.";
RT Mol. Biol. Rep. 23:205-210(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SVJ;
RA Kaufman K.M., Farris A.D., Gross J.K., Harley J.B.;
RT "Characterization of the murine 60 kD Ro gene: genomic sequence,
RT organization, and chromosomal localization.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 82-538 FROM N.A.
RA Scofield R.H., Kurien B.T., Kaufman K.M., James J.A., Baber U.,
RA Harley J.B.;
RT "Immunization of mice with human 60 kD Ro peptides results in epitope
RT spreading if the peptides are highly homologous between man and
RT mouse.";
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: RNA-BINDING PROTEIN THAT BINDS TO SEVERAL SMALL
CC CYTOPLASMIC RNA MOLECULES KNOWN AS Y RNAS. MAY STABILIZE THESE
CC RNAS FROM DEGRADATION.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: HIGHEST IN BRAIN, FOLLOWED BY LUNG, MUSCLE,
CC KIDNEY AND HEART. LOWER LEVELS ARE FOUND IN TESTIS, LIVER AND
CC SPLEEN.
CC -!- SIMILARITY: BELONGS TO THE RO 60 KDA FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U66843; AAC53142.1; -.
DR EMBL; AF065398; AAF19049.1; -.
DR EMBL; AF042139; AAC15667.1; -.
DR MGD; MGI:106652; Ssa2.
KW Ribonucleoprotein; RNA-binding.
FT CONFLICT 32 458 RL -> V (IN REF. 2).
FT CONFLICT 458 458 D -> G (IN REF. 2).

FT CONFLICT 465 465 V -> I (IN REF. 2).
SQ SEQUENCE 538 AA: 60123 MW: 7850DA35D1726BDA CRC64;

Query Match 65.5%; Score 36; DB 1; Length 538;
Best Local Similarity 50.0%; Pred. No. 46;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 KGGVVIATVVK 12
III: I I I
Db. 159 KGGMAVALVVT 170

RESULT 14
DCTS_RHOCA
ID DCTS_RHOCA STANDARD; PRT; 657 AA.
AC P37739;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DE C4-DICARBOXYLATE TRANSPORT SENSOR PROTEIN DCTS (EC 2.7.3.-).
GN DCTS.
OS Rhodobacter capsulatus (Rhodospseudomonas capsulata).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Rhodobacter.
OX NCBI_TaxID=1061;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B10;
RX MEDLINE=93204897; PubMed=8455557;
RA Hamblin M., Shaw J.G., Kelly D.J.;
RT "Sequence analysis and interposon mutagenesis of a sensor-kinase (DCTS) and response-regulator (Dctr) controlling synthesis of the high-affinity C4-dicarboxylate transport system in Rhodobacter capsulatus.";
RL Mol. Gen. Genet. 237:215-224(1993).
CC -1- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM DCTS/DCTR INVOLVED IN THE TRANSPORT OF C4-DICARBOXYLATES. DCTS FUNCTIONS AS A MEMBRANE-ASSOCIATED PROTEIN KINASE THAT PHOSPHORYLATES DCTR IN RESPONSE TO ENVIRONMENTAL SIGNALS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
CC -1- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE KINASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; X64733; CAA45999.1; -.
CC PIR; S30288; S30288.
CC InterPro; IPR000014; -.
CC InterPro; IPR000410; -.
CC Pfam; PF00785; PAC; 1.
CC Pfam; PF00989; PAS; 1.
CC Pfam; PF00512; signal; 1.
CC PRINTS; PR00344; BCTRSENSOR.
CC Sensory transduction; Transferrase; Kinase; Phosphorylation;
KW Transmembrane; Inner membrane.
FT DOMAIN 1 26 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 27 51 POTENTIAL.
FT DOMAIN 52 252 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 253 273 POTENTIAL.
FT DOMAIN 274 657 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 407 422 INTER-DOMAIN LINKER (POTENTIAL).
FT MOD_RES 440 440 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 657 AA: 70142 MW: 139D2CFC7CFFA69B CRC64;

Query Match 65.5%; Score 36; DB 1; Length 657;
Best Local Similarity 87.5%; Pred. No. 55;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGVVIATV 9
IIII:III
Db. 175 GGVVATV 182

RESULT 15
PECL_MOUSE
ID PECL_MOUSE STANDARD; PRT; 727 AA.
AC Q08481;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DE PLATELET ENDOTHELIAL CELL ADHESION MOLECULE PRECURSOR (PECAM-1)
DE (CD31 ANTIGEN).
GN PECAM1 OR PECAM-1 OR PECAM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=Lung;
RX MEDLINE=93256179; PubMed=8516303;
RA Xie Y., Muller W.A.;
RT "Molecular cloning and adhesive properties of murine platelet/endothelial cell adhesion molecule 1.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:5569-5573(1993).
RN [2]
RP SEQUENCE OF 18-26.
RC TISSUE=Heart;
RX MEDLINE=93015639; PubMed=1415479;
RA Bogen S.A., Baldwin H.S., Watkins S.C., Albelda S.M., Abbas A.K.;
RT "Association of murine CD31 with transmemigrating lymphocytes following antigenic stimulation.";
RL Am. J. Pathol. 141:843-854(1992).
CC -1- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE EXPRESSED ON PLATELETS AND AT ENDOTHELIAL CELL INTERCELLULAR JUNCTIONS.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; L06039; AAA16230.1; -.
CC MGD; MGI:97537; Pecam.
CC InterPro; IPR003006; -.
CC Pfam; PF00447; Ig; 3.
CC Immunoglobulin domain; Glycoprotein; Cell adhesion; Transmembrane;
KW Phosphorylation; Signal.
FT SIGNAL 1 17
FT CHAIN 18 727 PLATELET ENDOTHELIAL CELL ADHESION MOLECULE.
FT DOMAIN 18 590 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 591 609 POTENTIAL.
FT DOMAIN 610 727 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 40 106 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 135 203 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 238 300 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 329 382 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 413 472 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 505 568 IG-LIKE C2-TYPE DOMAIN.
FT DISULFID 47 99 POTENTIAL.
FT DISULFID 142 195 POTENTIAL.
FT DISULFID 245 293 POTENTIAL.

FT	DISULFID	336	375	POTENTIAL.
FT	DISULFID	420	465	POTENTIAL.
FT	DISULFID	512	561	POTENTIAL.
FT	CARBOHYD	74	74	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	141	141	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	309	309	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	345	345	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	360	360	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	424	424	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	540	540	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	MOD_RES	702	702	PHOSPHORYLATION (POTENTIAL).
FT	CONFLICT	18	18	E -> T (IN REF. 2).
SQ	SEQUENCE	727 AA; 81262 MW; 34C04752D199BAA5		CRG64;

Query Match 65.5%; Score 36; DB 1; Length 727;
 Best Local Similarity 88.9%; Pred. No. 60;
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 GVVIATVIV 11
 |||||:|
 Db 599 GVVIATLIV 607

Search completed: September 6, 2001, 16:51:07
 Job time: 811 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:45:49 ; Search time 134.15 Seconds
(without alignments)
6.814 Million cell updates/sec

Title: US-09-603-713-7
Perfect score: 55
Sequence: 1 KGGVVIATVIVK 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	81.8	57	2 E60045	Alzheimer's diseases
2	45	81.8	57	2 F60045	Alzheimer's diseases
3	45	81.8	57	2 G60045	Alzheimer's diseases
4	45	81.8	57	2 D60045	Alzheimer's diseases
5	45	81.8	57	2 A60045	Alzheimer's diseases
6	45	81.8	57	2 B60045	Alzheimer's diseases
7	45	81.8	82	2 P00438	Alzheimer's diseases
8	45	81.8	695	1 A49795	Alzheimer's diseases
9	45	81.8	695	2 A27485	Alzheimer's diseases
10	45	81.8	695	2 S00550	Alzheimer's diseases
11	45	81.8	747	2 Q07773	Alzheimer's diseases
12	45	81.8	770	1 QRRHUA	Alzheimer's diseases
13	39	70.9	438	2 F82944	GTP-binding protein
14	38	69.1	155	2 F84214	hypothetical protei
15	38	69.1	430	2 T35676	probable ABC-type
16	37	67.3	147	2 A69198	hypothetical prote
17	37	67.3	1196	2 H85061	hypothetical prote
18	36	65.5	427	2 C82221	conserved hypothet
19	36	65.5	536	2 J60022	flagellar basal-bo
20	36	65.5	657	2 S30288	C4-dicarboxylate s
21	35	63.6	217	2 F86343	hypothetical prote
22	35	63.6	236	2 C69060	hypothetical prote
23	35	63.6	298	2 F81301	probable integral
24	35	63.6	300	2 D84459	probable catlon tr
25	35	63.6	339	2 T34925	ABC transporter in
26	35	63.6	729	2 J01898	probable RNA-direc
27	35	63.6	740	2 S61568	probable membrane
28	35	63.6	1512	2 T14883	hypothetical prote
29	35	63.6	1819	2 A1928	cag island protein

cag pathogenicity
ABC transporter, p
hypothetical prote
hypothetical prote
probable mitochond
phytoalexin - Scot
serine protease do
MRK1 protein - yea
flagellar basal-bo
hypothetical prote
phosphate transpor
Na+/K+-exchanging
probable RND efflu
host cell factor C
genome polyprotein
hypothetical prote

ALIGNMENTS

RESULT 1

E60045
Alzheimer's disease amyloid beta/A4 protein precursor - sheep (fragment)
C:Species: Ovis sp. (sheep)
C:Date: 01-Dec-1992 #sequence_revision 01-Dec-1992 #text_change 28-Jul-1995
C:Accession: E60045
R:Johnstone, E.M.; Chaney, M.O.; Norris, F.H.; Pascual, R.; Little, S.P.
Brain Res. Mol. Brain Res. 10, 299-305, 1991
A:Title: Conservation of the sequence of the Alzheimer's disease amyloid peptide in d
A:Reference number: A60045; MUID:92017079
A:Accession: E60045
A:Molecule type: mRNA
A:Residues: 1-57 <JOH>

A:Cross-references: EMBL:X56130
C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinas
C:Keywords: alternative splicing; Alzheimer's disease; amyloid; brain

Query Match 81.8%; Score 45; DB 2; Length 57;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGVVIATVIV 11
Db 42 GGVVIATVIV 51

RESULT 2

F60045
Alzheimer's disease amyloid beta/A4 protein precursor - pig (fragment)
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 01-Dec-1992 #sequence_revision 01-Dec-1992 #text_change 13-Aug-1999
C:Accession: F60045
R:Johnstone, E.M.; Chaney, M.O.; Norris, F.H.; Pascual, R.; Little, S.P.
Brain Res. Mol. Brain Res. 10, 299-305, 1991
A:Title: Conservation of the sequence of the Alzheimer's disease amyloid peptide in d
A:Reference number: A60045; MUID:92017079
A:Accession: F60045
A:Molecule type: mRNA
A:Residues: 1-57 <JOH>

A:Cross-references: EMBL:X56127; NID:g1895; PIDN:CAA39592.1; PID:g1896
C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinas
C:Keywords: alternative splicing; Alzheimer's disease; amyloid; brain

Query Match 81.8%; Score 45; DB 2; Length 57;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGVVIATVIV 11
Db 42 GGVVIATVIV 51

Query Match	81.8%;	Score 45;	DB 2;	Length 57;	
Best Local Similarity	100.0%;	Pred. No. 0.25;			
Matches 10;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	2 GGVVIATVIV 11				
DB	42 GGVVIATVIV 51				
RESULT	6				
B60045	Alzheimer's disease amyloid beta/A4 protein precursor - polar bear (fragment)				
C:Species:	Ursus maritimus (polar bear)				
C:Date:	01-Dec-1992 #sequence_revision 01-Dec-1992 #text_change 13-Aug-1999				
C:Accession:	B60045				
R:Johnstone, E.M.;	Chaney, M.O.;			Norris, F.H.;	
Brain Res. Mol. Brain Res. 10, 299-305, 1991					
A:Title:	Conservation of the sequence of the Alzheimer's disease amyloid peptide in d				
A:Reference number:	A60045; MUID:92017079				
A:Accession:	B60045				
A:Molecule type:	mRNA				
A:Residues:	1-57 <JOH>				
A:Cross-references:	EMBL:X56128; NID:92165; PIDN:CAA39593.1; PID:G2166				
C:Superfamily:	Alzheimer's disease amyloid beta protein; animal kunitz-type proteinas				
C:Keywords:	alternative splicing; Alzheimer's disease; amyloid; brain				
Query Match	81.8%;	Score 45;	DB 2;	Length 57;	
Best Local Similarity	100.0%;	Pred. No. 0.25;			
Matches 10;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	2 GGVVIATVIV 11				
DB	42 GGVVIATVIV 51				
RESULT	7				
PQ0438	Alzheimer's disease amyloid A4 protein precursor - rabbit (fragment)				
C:Species:	Oryctolagus cuniculus (domestic rabbit)				
C:Date:	30-Sep-1993 #sequence_revision 19-Oct-1995 #text_change 19-Oct-1995				
C:Accession:	PQ0438; C60045				
R:Davidson, J.S.;	West, R.L.;			Kotikalapudi, P.;	
Biochem. Biophys. Res. Commun. 188, 905-911, 1992					
A:Title:	Sequence and methylation in the beta/A4 region of the rabbit amyloid precurs				
A:Reference number:	PQ0438; MUID:93075180				
A:Accession:	PQ0438				
A:Molecule type:	DNA				
A:Residues:	1-82 <DNA>				
A:Cross-references:	GB:M83657				
R:Johnstone, E.M.;	Chaney, M.O.;			Norris, F.H.;	
Brain Res. Mol. Brain Res. 10, 299-305, 1991					
A:Title:	Conservation of the sequence of the Alzheimer's disease amyloid peptide in d				
A:Reference number:	A60045; MUID:92017079				
A:Accession:	C60045				
A:Molecule type:	mRNA				
A:Residues:	12-68 <JOH>				
A:Cross-references:	EMBL:X56129				
C:Superfamily:	Alzheimer's disease amyloid beta protein; animal kunitz-type proteinas				
C:Keywords:	alternative splicing; Alzheimer's disease; amyloid; Down's syndrome				
Query Match	81.8%;	Score 45;	DB 2;	Length 82;	
Best Local Similarity	100.0%;	Pred. No. 0.34;			
Matches 10;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	2 GGVVIATVIV 11				
DB	53 GGVVIATVIV 62				
RESULT	8				

```
A49795
Alzheimer's disease amyloid beta protein precursor - crab-eating macaque
C:Species: Macaca fascicularis (crab-eating macaque)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A49795
R:Podlasky, M.B.; Tolan, D.R.; Selkoe, D.J.
Am. J. Pathol. 138, 1423-1435, 1991
A:Title: Homology of the amyloid beta protein precursor in monkey and human supports a
A:Reference number: A49795; MUID:91273117
A:Accession: A49795
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-695 <PDB>
A:Cross-references: GB:M58727; NID:g342062; PIDN:AAA36829.1; PID:g342063
C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinase
C:Keywords: alternative splicing

Query Match 81.8%; Score 45; DB 1; Length 695;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GGVVATVIV 11
Db 633 GGVVATVIV 642
|||||

RESULT 9
A27485
Alzheimer's disease amyloid beta/A4 protein homolog precursor - mouse
A:Alternate names: proteinase nexin II
C:Species: Mus musculus (house mouse)
C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 13-Aug-1999
C:Accession: A27485; S19727; I49485
R:Yamada, T.; Sakaki, H.; Furuya, H.; Miyata, T.; Goto, I.; Sakaki, Y.
Biochem. Biophys. Res. Commun. 149, 665-671, 1987
A:Title: Complementary DNA for the mouse homolog of the human amyloid beta protein precu
A:Reference number: A27485; MUID:88106489
A:Accession: A27485
A:Molecule type: mRNA
A:Residues: 1-695 <YAM>
A:Cross-references: GB:M18373; NID:g191568; PIDN:AAA37139.1; PID:g305085
A:Experimental source: brain
R:de Strooper, B.; van Leuven, F.; van den Berghe, H.
Biochim. Biophys. Acta 1129, 141-143, 1991
A:Title: The amyloid beta protein precursor or proteinase nexin II from mouse is closer
A:Reference number: S19727; MUID:92096458
A:Accession: S19727
A:Molecule type: mRNA
A:Residues: 1-210, 'G', 212-220, 'S', 222-396, 'A', 398-402, 'T', 404-448, 'A', 450-695 <STR>
A:Cross-references: EMBL:X59379
R:Izumi, R.; Yamada, T.; Yoshikai, S.; Sasaki, H.; Hattori, M.; Sakaki, Y.
Gene 112, 189-195, 1992
A:Title: Positive and negative regulatory elements for the expression of the Alzheimer's
A:Reference number: I49485; MUID:92209998
A:Accession: I49485
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-19 <RES>
A:Cross-references: GB:D10603; NID:g220328; PIDN:BAA01456.1; PID:g220329
C:Genetics:
A:Map position: 16C3
C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinase
C:Keywords: alternative splicing; amyloid; transmembrane protein

Query Match 81.8%; Score 45; DB 2; Length 695;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GGVVATVIV 11
Db 633 GGVVATVIV 642
|||||
```

```
RESULT 10
S00550
Alzheimer's disease amyloid beta protein precursor - rat
A:Alternate names: beta-A4 amyloid protein
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 13-Aug-1999
C:Accession: S00550; A41245; A39820; S46251
R:Shivers, B.D.; Hilbich, C.; Multhaup, G.; Salbaum, M.; Beyreuther, K.; Seeburg, P.H.
EMBO J. 7, 1365-1370, 1988
A:Title: Alzheimer's disease amyloidogenic glycoprotein: expression pattern in rat br
A:Reference number: S00550; MUID:88312583
A:Accession: S00550
A:Molecule type: mRNA
A:Residues: 1-635 <SHI>
A:Cross-references: EMBL:X07648; NID:g55616; PIDN:CAA30488.1; PID:g55617
R:Schubert, D.; Schroeder, R.; LaCorbiere, M.; Saitoh, T.; Cole, G.
Science 241, 223-226, 1988
A:Title: Amyloid beta protein precursor is possibly a heparan sulfate proteoglycan co
A:Reference number: A41245; MUID:88264430
A:Accession: A41245
A:Molecule type: protein
A:Residues: 18-17, 'X', 39-40, 'X', 42-44 <SCH>
A:Note: evidence for heparan sulfate attachment
R:Hesse, L.; Behner, D.; Masters, C.L.; Multhaup, G.
FEBS Lett. 349, 109-116, 1994
A:Title: The beta-A4 amyloid precursor protein binding to copper.
A:Reference number: S46251; MUID:94320627
A:Contents: annotation; copper binding sites
A:Note: rat peptides were isolated but not sequenced
R:Potempska, A.; Styles, J.; Mehta, P.; Kim, K.S.; Miller, D.L.
J. Biol. Chem. 266, 8464-8469, 1991
A:Title: Purification and tissue level of the beta-amyloid peptide precursor of rat b
A:Reference number: A39820; MUID:91217087
A:Accession: A39820
A>Status: preliminary
A:Molecule type: protein
A:Residues: 18-12 <POT>
A:Experimental source: brain
C:Comment: Deposition of amyloid protein as neurofibrillary tangles and/or plaques is
C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinas
C:Keywords: alternative splicing; amyloid; glycoprotein; transmembrane protein
F:625-648/Domain: transmembrane #status predicted <TMM>

Query Match 81.8%; Score 45; DB 2; Length 695;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GGVVATVIV 11
Db 633 GGVVATVIV 642
|||||

RESULT 11
JH0773
Alzheimer's disease amyloid beta protein precursor - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 10-Jun-1993 #sequence_revision 10-Jun-1993 #text_change 13-Aug-1999
C:Accession: JH0773
R:Okado, H.; Okamoto, H.
Biochem. Biophys. Res. Commun. 189, 1561-1568, 1992
A:Title: A Xenopus homologue of the human beta-amyloid precursor protein: development
A:Reference number: JH0773; MUID:93129227
A:Accession: JH0773
A:Molecule type: mRNA
A:Residues: 1-747 <OKA>
A:Cross-references: GB:S52417; NID:g263150; PIDN:AAB24853.1; PID:g263151
A:Experimental source: larva
C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinas
C:Keywords: alternative splicing; amyloid
F:287-337/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>
```

Query Match 81.98; Score 45; DB 2; Length 747;
 Best Local Similarity 100.0%; Pred. No. 2.5;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGVIATVIV 11
 |||||
 Db 695 GGVIATVIV 694

RESULT 12

ORHUA4
 Alzheimer's disease amyloid beta protein precursor [validated] - human
 N:Alternate names: Alzheimer's disease amyloid A4 protein; coagulation factor xia inhibitor
 N:Contains: amyloid beta protein long, plaque form; amyloid beta protein short, vascular
 protein precursor splice form APP(770)
 C:Species: Homo sapiens (man)
 C:Date: 30-Jun-1987 #sequence_revision 28-Jul-1995 #text_change 15-Sep-2000
 C:Accession: S02260; S05194; A32277; A33260; A35486; I39452; I39453; I39455; S3
 4668; A28583; A29302; A60805; J10038; S06121; A60355; A59011; A38384; S29076; S3
 R:Lemaire, H.G.; Salbaum, J.M.; Multhaup, G.; Kang, J.; Bayney, R.M.; Unterbeck, A.; Bey
 Nucleic Acids Res. 17, 517-522, 1989
 A:Title: The PreA4(695) precursor protein of Alzheimer's disease A4 amyloid is encoded b
 A:Reference number: S02260; MUID:89128427
 A:Accession: S02260
 A:Molecule type: DNA
 A:Residues: 1-288, 'V', 365-770 <LEM1>
 A:Cross-references: EMBL:X13466
 A:Note: alternative splice form APP(695)
 A:Title: alternative splice form APP(695)
 R:Lemaire, H.G.
 submitted to the EMBL Data Library, November 1988
 A:Reference number: S05194
 A:Accession: S05194
 A:Molecule type: DNA
 A:Residues: 1-14, 'VW', 17-288, 'V', 365-770 <LEM2>
 A:Cross-references: EMBL:X13466; NID:G35598; PIDN:CAA31830.1; PID:g871360
 A:Note: alternative splice form APP(695)
 R:La Fauci, G.; Lahiri, D.K.; Salton, S.R.J.; Robakis, N.K.
 Biochem. Biophys. Res. Commun. 159, 297-304, 1989
 A:Title: Characterization of the 5'-end region and the first two exons of the beta-prote
 A:Reference number: A32277; MUID:89165870
 A:Accession: A32277
 A:Molecule type: DNA
 A:Residues: 1-75 <LAF>
 A:Cross-references: GB:M24546; GB:M24547; NID:G341202; PIDN:AAC13654.1; PID:g516074
 R:Johnstone, E.M.; Chaney, M.O.; Moore, R.E.; Ward, K.E.; Norris, F.H.; Little, S.P.
 Biochem. Biophys. Res. Commun. 163, 1248-1255, 1989
 A:Title: Alzheimer's disease amyloid peptide is encoded by two exons and shows similarit
 A:Reference number: A33260; MUID:89392030
 A:Accession: A33260
 A:Molecule type: DNA
 A:Residues: 656-737 <JOH>
 A:Cross-references: GB:M29270; NID:g178863; PIDN:AAA51768.1; PID:g178865
 R:Prelli, F.; Levy, E.; Van Duinen, S.G.; Bots, G.T.A.M.; Luyendijk, W.; Frangione, B.
 Biochem. Biophys. Res. Commun. 170, 301-307, 1990
 A:Title: Expression of a normal and variant Alzheimer's beta-protein gene in amyloid of
 A:Reference number: A35486; MUID:90321244
 A:Accession: A35486
 A:Molecule type: DNA
 A:Residues: 672-710 <PREL>
 A:Note: 693-Gln was found in DNA isolated from HCHWA-D patients
 R:Yoshikai, S.I.; Sasaki, H.; Doh-ura, K.; Furuya, H.; Sasaki, Y.
 Gene 87, 257-263, 1990
 A:Title: Genomic organization of the human amyloid beta-protein precursor gene.
 A:Reference number: I39451; MUID:90236318
 A:Accession: I39452
 A>Status: nucleic acid sequence not shown; translation not shown; translated from GB/EMB
 A:Molecule type: DNA
 A:Residues: 1-770 <YOSI>
 A:Cross-references: GB:M33112; NID:g178613; PIDN:AAB59502.1; PID:g178616
 A:Accession: I39451
 A>Status: nucleic acid sequence not shown; translation not shown; translated from GB/EMB

A:Molecule type: DNA
 A:Residues: 1-530, 'OWLMPVTPAFWEAKVGR' <YOS2>
 A:Cross-references: GB:M34875; NID:g178608; PIDN:AAB59501.1; PID:g178615
 R:Yoshikai, S.I.; Sasaki, H.; Doh-ura, K.; Furuya, H.; Sasaki, Y.
 Gene 102, 291-292, 1991
 A:Reference number: A59020; MUID:91340168
 A:Contents: annotation; erratum
 A:Note: revised physical map for reference I39451
 R:Levy, E.; Carman, M.D.; Fernandez-Madrid, I.J.; Power, M.D.; Lieberburg, I.; van Du
 Science 248, 1124-1126, 1990
 A:Title: Mutation of the Alzheimer's disease amyloid gene in hereditary cerebral hemo
 A:Reference number: I39453; MUID:90260663
 A:Accession: I39453
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 656-737 <LEV>
 A:Cross-references: GB:M37896; NID:g178618; PIDN:AAA51727.1; PID:g178620
 A:Note: a mutation with 693-Gln is presented
 R:Murrell, J.; Farlow, M.; Ghetti, B.; Benson, M.D.
 Science 254, 97-99, 1991
 A:Title: A mutation in the amyloid precursor protein associated with hereditary Alzhe
 A:Reference number: I59562; MUID:92022553
 A:Accession: I59562
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 689-716, 'F', 718-737 <MUR>
 A:Cross-references: GB:S57665; NID:G236720; PIDN:AAB19991.1; PID:G236721
 R:Kamino, K.; Orr, H.T.; Payami, H.; Wijsman, E.M.; Alonso, M.E.; Pulst, S.M.; Anders
 arakis, S.E.; Korenberg, J.R.; Sharma, V.; Kukull, W.; Larson, E.; Heston, L.L.; Mart
 Am. J. Hum. Genet. 51, 998-1014, 1992
 A:Title: Linkage and mutational analysis of familial Alzheimer disease kindreds for t
 A:Reference number: A44017; MUID:93035397
 A:Accession: A44017
 A:Molecule type: DNA
 A:Residues: 687-692, 'G', 694-718 <KAM1>
 A:Cross-references: GB:S45135; NID:G257377; PIDN:AAB23645.1; PID:G257378
 A:Experimental source: familial Alzheimer disease family SB
 A:Note: sequence extracted from NCBI backbone (NCBIP:115374)
 A:Accession: B44017
 A:Molecule type: DNA
 A:Residues: 687-718 <KAM2>
 A:Cross-references: GB:S45136; NID:G257379; PIDN:AAB23646.1; PID:G257380
 A:Experimental source: familial Alzheimer disease family LIT
 A:Note: sequence extracted from NCBI backbone (NCBIP:115376)
 A:Note: this sequence has a silent mutation
 R:Kang, J.; Lemaire, H.G.; Unterbeck, A.; Salbaum, J.M.; Masters, C.L.; Grzeschik, K.
 Nature 325, 733-736, 1987
 A:Title: The precursor of Alzheimer's disease amyloid A4 protein resembles a cell-sur
 A:Reference number: A03134; MUID:87144572
 A:Accession: A03134
 A:Molecule type: mRNA
 A:Residues: 1-288, 'V', 365-770 <KAN>
 A:Cross-references: GB:Y00264; NID:G28525; PIDN:CAA68374.1; PID:G28526
 A:Note: alternative splice form APP(695)
 R:Robakis, N.K.; Ramakrishna, N.; Wolfe, G.; Wisniewski, H.M.
 Proc. Natl. Acad. Sci. U.S.A. 84, 4190-4194, 1987
 A:Title: Molecular cloning and characterization of a cDNA encoding the cerebrovascula
 A:Reference number: A29030; MUID:87231971
 A:Accession: A29030
 A:Molecule type: mRNA
 A:Residues: 284-288, 'V', 365-646, 'E', 648-770 <ROB>
 A:Cross-references: GB:M16765; NID:G178539; PIDN:AAA51722.1; PID:g178540
 A:Note: the authors translated the codon GAG for residue 647 as Asp
 R:Goldgaber, D.; Lerman, M.I.; McBride, O.W.; Saffioti, U.; Gajdusek, D.C.
 Science 235, 877-880, 1987
 A:Title: Characterization and chromosomal localization of a cDNA encoding brain amyl
 A:Reference number: A47584; MUID:87120328
 A:Accession: A47584
 A:Molecule type: mRNA
 A:Residues: 674-736, 'S', 758-770 <GOL>
 A:Cross-references: GB:M15533; NID:g178706; PIDN:AAA35540.1; PID:g178707
 A:Experimental source: brain
 R:Tanzi, R.E.; Gusella, J.F.; Watkins, P.C.; Bruns, G.A.P.; St George-Hyslop, P.; Van

Science 235, 880-884, 1987
A:Title: Amyloid beta protein gene: cDNA, mRNA distribution, and genetic linkage near the
A:Reference number: A47585; MUID:87120329
A:Accession: A47585
A:Molecule type: mRNA
A:Residues: 674-703 <TANI>
A:Cross-references: GB:M15532; NID:g177957; PIDN:AAA51564.1; PID:g177958
R:Dyrks, T.; Weidemann, A.; Multhaup, G.; Salbaum, J.M.; Lemaire, H.G.; Kang, J.; Muelle
EMBO J. 7, 949-957, 1988
A:Title: Identification, transmembrane orientation and biogenesis of the amyloid A4 pre
A:Reference number: S02638; MUID:88296437
A:Accession: S02638
A:Molecule type: mRNA
A:Residues: 672-678 <DVR>
R:Tanzi, R.E.; McClatchey, A.I.; Lamperti, E.D.; Villa-Komaroff, L.; Gusella, J.F.; Neve
Nature 331, 528-530, 1988
A:Title: Protease inhibitor domain encoded by an amyloid protein precursor mRNA associat
A:Reference number: S00707; MUID:88122640
A:Accession: S00707
A:Molecule type: mRNA
A:Residues: 286-344, 'I', 365-366 <TAN2>
A:Cross-references: EMBL:X06982; NID:g28817; PIDN:CAA30042.1; PID:g929612
A:Experimental source: promyelocytic leukemia cell line HL60
A:Note: alternative splice form APP(751)
R:Ponte, P.; Gonzalez-Dewhitt, P.; Schilling, J.; Miller, J.; Hsu, D.; Greenberg, B.; Da
Nature 331, 530-532, 1988
A:Title: A new A4 amyloid mRNA contains a domain homologous to serine proteinase inhibit
A:Reference number: S00925; MUID:88122639
A:Accession: S00925
A:Molecule type: mRNA
A:Residues: 1-344, 'I', 365-770 <PO2>
A:Cross-references: GB:X06983; EMBL:X06989; EMBL:Y00297; NID:g28720; PIDN:CAA30050.1; PID:g28721
A:Note: alternative splice form APP(751)
R:Kitaguchi, N.; Takahashi, Y.; Tokushima, Y.; Shiojiri, S.; Ito, H.
Nature 331, 530-532, 1988
A:Title: Novel precursor of Alzheimer's disease amyloid protein shows protease inhibitor
A:Reference number: A38949; MUID:88122641
A:Accession: A38949
A:Molecule type: mRNA
A:Residues: 287-367 <KIT>
A:Cross-references: GB:X06981; NID:g28816; PIDN:CAA30041.1; PID:g929611
A:Experimental source: glioblastoma cell line
A:Note: alternative splice form APP(770)
R:Vitek, M.P.; Rasool, C.G.; de Sauvage, F.; Vitek, S.M.; Bartus, R.T.; Beer, B.; Ashton
Brain Res. Mol. Brain Res. 4, 121-131, 1988
A:Title: Absence of mutation in the beta-amyloid cDNAs cloned from the brains of three P
A:Reference number: A30320
A:Accession: A30320
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 284-288, 'V', 365-770 <VIT1>
A:Accession: B30320
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 122-288, 'V', 365-770 <VIT2>
A:Accession: C30320
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 606-770 <VIT3>
R:Zain, S.B.; Salim, M.; Chou, W.G.; Sajdel-Sulkowska, E.M.; Majocha, R.E.; Marotta, C.A
Proc. Natl. Acad. Sci. U.S.A. 85, 929-933, 1988
A:Title: Molecular cloning of amyloid cDNA derived from mRNA of the Alzheimer disease br
A:Reference number: A31087; MUID:88124954
A:Accession: A31087
A:Molecule type: mRNA
A:Residues: 507-770 <ZAI>
A:Cross-references: GB:M18734; NID:g178572; PIDN:AAA51726.1; PID:g178573
A:Note: the authors translated the codon GAA for residue 599 as Gly, ACC for residue 603
8 as Val, GTG for residue 609 as Asn, AAT for residue 610 as Gly, and GGT for residue 65
A:Note: the cited genbank accession number, J03594, is not in release 101.0
R:Masters, C.L.; Multhaup, G.; Simms, G.; Potgiesser, J.; Martins, R.N.; Beyreuther, K.

Query Match

81.8%; Score 45; DB 1; Length 770;

Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2 GGVVIATVIV 11
Db 708 GGVVIATVIV 717
RESULT 13
F82944
GTP-binding protein in thiophene and furan oxidation U0018 [Imported] - Ureaplasma ur
C:Species: Ureaplasma urealyticum
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.
submitted to GenBank, February 2000
A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a
A:Reference number: A82870
A:Accession: F82944
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-438 <GLA>
A:Cross-references: GB:AE002101; GB:AF222894; NID:g6898957; PIDN:AAF30423.1; GSPDB:GN
A:Experimental source: serovar 3; biovar 1
C:Genetics:
A:Gene: thdF; U0018
A:Genetic code: SGC3
C:Superfamily: thiophen / furan oxidation protein; translation elongation factor Tu h

Query Match 70.9%; Score 39; DB 2; Length 438;
Best Local Similarity 54.5%; Pred. No. 19;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GGVVIATVIVK 12
Db 83 GGIVANLIIR 93

RESULT 14

F84214
hypothetical protein Vng0564h [Imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Love, T.M.;
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483
A:Accession: F84214
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-155 <STO>
A:Cross-references: GB:AE004437; NID:g10580162; PIDN:AAG19082.1; GSPDB:GN00138
C:Genetics:
A:Gene: VNG0564H

Query Match 69.1%; Score 38; DB 2; Length 155;
Best Local Similarity 54.5%; Pred. No. 11;
Matches 6; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KGGVVIATVIV 11
Db 50 QGGVIVATVNL 60

RESULT 15

T35676
probable ABC-type transmembrane transport protein - Streptomyces coelicolor

C:Species: Streptomyces coelicolor
 C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
 C:Accession: T35676
 R:Murphy, L.; Harris, D.; Parkhill, J.; Barrrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, November 1997
 A:Reference number: Z21586
 A:Accession: T35676
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-430 <MUR>
 A:Cross-references: EMBL:AL009199; PIDN:CAAL5788.1; GSPDB:GN00070; SCOEDB:SC7B7.08
 A:Experimental source: strain A3(2)
 C:Genetics:
 A:Gene: SCOEDB:SC7B7.08

Query Match	69.1%	Score 38;	DB 2;	Length 430;
Best Local Similarity	60.0%	Pred. No. 28;		
Matches	6;	Conservative 4;	Mismatches 0;	Indels 0;
			Gaps 0;	
Qy	2	GGVVIATVIV 11		
		:		
Db	407	GGVLLATVVI 416		

Search completed: September 6, 2001, 16:45:50
 Job time: 494 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:39:35 ; Search time 113.12 Seconds
(without alignments)
2.184 Million cell updates/sec

Title: US-09-603-713-7
Perfect score: 55
Sequence: 1 KGGVVIATVIVK 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/1aa/5A-COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B-COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A-COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B-COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PTCUS-COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	81.8	23	3	US-08-802-981-117
2	45	81.8	24	1	US-08-371-930-8
3	45	81.8	24	5	PCT-US94-01712-8
4	45	81.8	45	1	US-08-462-859A-5
5	45	81.8	45	1	US-08-123-659A-5
6	45	81.8	45	1	US-08-464-247A-5
7	45	81.8	45	1	US-08-464-248A-5
8	45	81.8	47	2	US-08-609-090-10
9	45	81.8	49	1	US-08-123-702-45
10	45	81.8	52	2	US-08-609-090-11
11	45	81.8	53	4	US-09-173-887-5
12	45	81.8	59	1	US-08-484-969-3
13	45	81.8	59	1	US-08-472-627-3
14	45	81.8	59	1	US-08-388-463-3
15	45	81.8	63	1	US-08-462-859A-3
16	45	81.8	63	1	US-08-462-859A-4
17	45	81.8	63	1	US-08-123-659A-3
18	45	81.8	63	1	US-08-123-659A-4
19	45	81.8	63	1	US-08-464-247A-3
20	45	81.8	63	1	US-08-464-247A-4
21	45	81.8	63	1	US-08-464-248A-3
22	45	81.8	63	1	US-08-464-248A-4
23	45	81.8	97	6	5187153-8
24	45	81.8	97	6	5220013-8
25	45	81.8	97	6	5223482-8
26	45	81.8	99	2	US-08-422-333-3
27	45	81.8	99	3	US-08-339-708A-4

28	45	81.8	99	3	US-08-339-708A-6
29	45	81.8	100	6	5187153-10
30	45	81.8	100	6	5220013-10
31	45	81.8	100	6	5223482-10
32	45	81.8	103	2	US-08-404-831-2
33	45	81.8	103	2	US-08-612-785B-2
34	45	81.8	103	2	US-08-475-579A-2
35	45	81.8	103	2	US-08-920-162A-2
36	45	81.8	103	3	US-08-339-708A-10
37	45	81.8	103	3	US-08-339-708A-12
38	45	81.8	105	2	US-08-729-345-1
39	45	81.8	108	6	5187153-14
40	45	81.8	108	6	5220013-18
41	45	81.8	108	6	5223482-16
42	45	81.8	117	2	US-08-729-345-3
43	45	81.8	264	1	US-07-990-893-5
44	45	81.8	487	1	US-08-462-859A-9
45	45	81.8	487	1	US-08-123-659A-9

ALIGNMENTS

RESULT 1
US-08-802-981-117
; Sequence 117, Application US/08802981
; Patent No. 6037137
; GENERAL INFORMATION:
; APPLICANT: Komoriya, Akira
; APPLICANT: Packard, Beverly S.
; TITLE OF INVENTION: Compositions for the Detection of Enzyme
; TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof
; NUMBER OF SEQUENCES: 231
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/802,981
; FILING DATE: 20-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunter, Tom
; REGISTRATION NUMBER: 38,498
; REFERENCE/DOCKET NUMBER: 016865-00030005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 117:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 23 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 3
; OTHER INFORMATION: /product= "Aib"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 4
; OTHER INFORMATION: /product= "Acp"
; FEATURE:
; NAME/KEY: Modified-site

RESULT 3
PCT-US94-01712-8
; Sequence 8, Application PC/TUS9401712
; GENERAL INFORMATION:
; APPLICANT: Nishimoto, Ikuo
; TITLE OF INVENTION: ALZHEIMER'S D

```
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 24
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-371-930-8
```

```

ZIP: 07/MAY/95
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,859A
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435

```

ATTORNEY/AGENT INFORMATION:
NAME: Barnhard, Elizabeth M.
REGISTRATION NUMBER: 31,088
REFERENCE/DOCKET NUMBER: 31,844-04
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201)831-3246
TELEFAX: (201)831-3305
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 45 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-462-859A-5

Query Match 81.8%; Score 45; DB 1; Length 45;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGVVIATVIV 11
Db 28 GGVVIATVIV 37
|||||

RESULT 5
US-08-123-659A-5
Sequence 5, Application US/08123659A
Patent No. 5656477
GENERAL INFORMATION:
APPLICANT: Jacobsen, J. S.
APPLICANT: Vitek, M. P.
TITLE OF INVENTION: No. 5656477el Amyloid Precursor and Method of
TITLE OF INVENTION: Using Same to Access Agents Which Down-Regulate Formation
TITLE OF INVENTION: of B-Amyloid Peptide
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Anne Rosenblum
STREET: 163 Delaware Avenue, Suite 212
CITY: Delmar
STATE: New York
COUNTRY: U.S.A.
ZIP: 12054

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/123,659A
FILING DATE: 20-SEP-1993
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Rosenblum, Anne M.
REGISTRATION NUMBER: 30,419
REFERENCE/DOCKET NUMBER: 31,844-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (518)475-0611
TELEFAX: (518)475-0619
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 45 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-123-659A-5

Query Match 81.8%; Score 45; DB 1; Length 45;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGVVIATVIV 11
Db 28 GGVVIATVIV 37
|||||

RESULT 6
US-08-464-247A-5
Sequence 5, Application US/08464247A
Patent No. 5693478
GENERAL INFORMATION:
APPLICANT: Jacobsen, J. S.
APPLICANT: Vitek, M. P.
TITLE OF INVENTION: No. 5693478el Amyloid Precursor and Method of
TITLE OF INVENTION: Using Same to Access Agents Which Down-Regulate Formation
TITLE OF INVENTION: of B-Amyloid Peptide
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: American Cyanamid Company
STREET: One Campus Drive
CITY: Parsippany
STATE: New Jersey
COUNTRY: United States
ZIP: 07054

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/464,247A
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Barnhard, Elizabeth M.
REGISTRATION NUMBER: 31,088
REFERENCE/DOCKET NUMBER: 31,844-03
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-683-2158
TELEFAX: 201-683-4117
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 45 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-464-247A-5

Query Match 81.8%; Score 45; DB 1; Length 45;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGVVIATVIV 11
Db 28 GGVVIATVIV 37
|||||

RESULT 7
US-08-464-248A-5
Sequence 5, Application US/08464248A
Patent No. 5703209
GENERAL INFORMATION:
APPLICANT: Jacobsen, J. S.
APPLICANT: Vitek, M. P.
TITLE OF INVENTION: No. 5703209el Amyloid Precursor and Method of
TITLE OF INVENTION: Using Same to Access Agents Which Down-Regulate Formation
TITLE OF INVENTION: of B-Amyloid Peptide
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: American Cyanamid Company
STREET: One Cyanamid Plaza
CITY: Wayne

us-09-603-713-7.ra

Fri Sep 7 10:58:46 2001

STATE: New Jersey
COUNTRY: United States
ZIP: 07470-8426
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/464,248A
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Barnhard, Elizabeth M.
REGISTRATION NUMBER: 31,088
REFERENCE/DOCKET NUMBER: 31,844-02
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201)831-3246
TELEFAX: (201)831-3305
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 45 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-464-248A-5

Query Match 81.8%; Score 45; DB 1; Length 45;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGVVIATVIV 11
| | | | | | | | | |
Db 28 GGVVIATVIV 37

RESULT 8
US-08-609-090-10
Sequence 10, Application US/08609090
Patent No. 5840838
GENERAL INFORMATION:
APPLICANT: HENSLEY, Kenneth
APPLICANT: BUTTERFIELD, D. A.
APPLICANT: CARNEY, John M.
APPLICANT: ARSENOV, Michael
TITLE OF INVENTION: A PROCESS FOR ENHANCING THE ACTIVITY OF
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: LOWE PRICE LEBLANC & BECKER
STREET: 99 Canal Center Plaza, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: USA
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/609,090
FILING DATE: 29-FEB-1996
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Kraus, Eric J.
REGISTRATION NUMBER: 36,190
REFERENCE/DOCKET NUMBER: 434-059
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-684-1111
TELEFAX: 703-684-1124

INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 47 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-609-090-10

Query Match 81.8%; Score 45; DB 2; Length 47;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGVVIATVIV 11
| | | | | | | | | |
Db 37 GGVVIATVIV 46

RESULT 9
US-08-123-702-45
Sequence 45, Application US/08123702
Patent No. 5604131
GENERAL INFORMATION:
APPLICANT: Wadsworth, Samuel
APPLICANT: Snyder, Benjamin
APPLICANT: Reddy, Vermuri, B.
APPLICANT: Wei, Chamer
TITLE OF INVENTION: A cDNA Genomic Hybrid Sequence Encoding APP770
PATENT NO. 5604131
TITLE OF INVENTION: Containing a Genomic DNA Insert of the KI and OX-2 Regions
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center
STREET: 1201 West Peachtree Street
CITY: Atlanta
STATE: GA
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/123,702
FILING DATE: 17-SEPT-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: TS1121
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404)-873-8794
TELEFAX: (404)-873-8795
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 49 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: mutation
LOCATION: 29
OTHER INFORMATION: "Val can be mutated to be Phe"
US-08-123-702-45

Query Match 81.8%; Score 45; DB 1; Length 49;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGVWATVIV 11
| | | | | | | |
DB 20 GGVWATVIV 29

RESULT 10
US-08-609-090-11
; Sequence 11, Application US/08609090
; Patent No. 5840838
; GENERAL INFORMATION:
; APPLICANT: HENSLEY, Kenneth
; APPLICANT: BUTTERFIELD, D. A.
; APPLICANT: CARNEY, John M.
; APPLICANT: AKSENOV, Michael
; TITLE OF INVENTION: A PROCESS FOR ENHANCING THE ACTIVITY OF
; TITLE OF INVENTION: AN OLIGOPEPTIDE OR POLYPEPTIDES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LOWE PRICE LEBLANC & BECKER
; STREET: 99 Canal Center Plaza, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/609,090
; FILING DATE: 29-FEB-1996
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Kraus, Eric J.
; REGISTRATION NUMBER: 36,190
; REFERENCE/DOCKET NUMBER: 434-059
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-684-1111
; TELEFAX: 703-684-1124
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 52 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-609-090-11

Query Match 81.8%; Score 45; DB 2; Length 52;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGVWATVIV 11
| | | | | | | |
DB 37 GGVWATVIV 46

RESULT 11
US-09-173-887-5
; Sequence 5, Application US/09173887
; Patent No. 6245884
; GENERAL INFORMATION:
; APPLICANT: Hook, Vivian Y.H.
; TITLE OF INVENTION: SECRETFASES RELATED TO AL2HEIMER'S DEMENTIA
; FILE REFERENCE: P-AS 3337
; CURRENT APPLICATION NUMBER: US/09/173.887
; CURRENT FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 53

; TYPE: PRT
; ORGANISM: mammalian
US-09-173-887-5

Query Match 81.8%; Score 45; DB 4; Length 53;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGVWATVIV 11
| | | | | | | |
DB 40 GGVWATVIV 49

RESULT 12
US-08-484-969-3
; Sequence 3, Application US/08484969
; Patent No. 5679531
; GENERAL INFORMATION:
; APPLICANT: Konig, Gerhard
; APPLICANT: Graham, Paul
; TITLE OF INVENTION: Monoclonal Antibody Specific for BA4
; TITLE OF INVENTION: Peptide
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Allegretti & Witcoff, Ltd.
; STREET: 10 South Wacker Drive Suite 3000
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,969
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: McDonnell, John J
; REGISTRATION NUMBER: 26,949
; REFERENCE/DOCKET NUMBER: 95,216
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-715-1000
; TELEFAX: 312-715-1234
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 59 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Cleavage-site
; LOCATION: 4..5
; OTHER INFORMATION: /label= Beta
; OTHER INFORMATION: /note= "Beta cleavage site in App"
; FEATURE:
; NAME/KEY: Cleavage-site
; LOCATION: 20..21
; OTHER INFORMATION: /label= Alpha
; OTHER INFORMATION: /note= "Alpha cleavage site in APP, residues 16/17
; OTHER INFORMATION: of BA4."
; FEATURE:
; NAME/KEY: Cleavage-site
; LOCATION: 46..47
; OTHER INFORMATION: /label= Gamma
; OTHER INFORMATION: /note= "Gamma cleavage site in App"
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 5..47

```
; OTHER INFORMATION: /label= BA4
; OTHER INFORMATION: /note= "BA4 peptide"
;
; FEATURE:
; NAME/KEY: Region
; LOCATION: 33..56
; OTHER INFORMATION: /label= Tm
; OTHER INFORMATION: /note= "Transmembrane region of App"
;
; FEATURE:
; NAME/KEY: Region
; LOCATION: 1..32
; OTHER INFORMATION: /label= Ex
; OTHER INFORMATION: /note= "N-terminal extracellular part of App"
;
US-08-484-969-3

Query Match      81.8%; Score 45; DB 1; Length 59;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  2 GGVVIATVIV 11
Db   41 GGVVIATVIV 50

RESULT 13
US-08-472-627-3
; Sequence 3, Application US/08472627
; Patent No. 5693753
; GENERAL INFORMATION:
; APPLICANT: Konig, Gerhard
; APPLICANT: Graham, Paul
; TITLE OF INVENTION: Monoclonal Antibody Specific for BA4
; TITLE OF INVENTION: Peptide
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Allegrretti & Witcoff, Ltd.
; STREET: 10 South Wacker Drive Suite 3000
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472.627
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: McDonnell, John J
; REGISTRATION NUMBER: 26,949
; REFERENCE/DOCKET NUMBER: 95,216
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-715-1000
; TELEFAX: 312-715-1234
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 59 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Cleavage-site
; LOCATION: 4..5
; OTHER INFORMATION: /label= Beta
; OTHER INFORMATION: /note= "Beta cleavage site in App"
;
; FEATURE:
; NAME/KEY: Cleavage-site
; LOCATION: 20..21
; OTHER INFORMATION: /label= Alpha
```

```
; OTHER INFORMATION: /note= "Alpha cleavage site in APP, residues 16/17
; OTHER INFORMATION: Of BA4."
;
; FEATURE:
; NAME/KEY: Cleavage-site
; LOCATION: 46..47
; OTHER INFORMATION: /label= Gamma
; OTHER INFORMATION: /note= "Gamma cleavage site in App"
;
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 5..47
; OTHER INFORMATION: /label= BA4
; OTHER INFORMATION: /note= "BA4 peptide"
;
; FEATURE:
; NAME/KEY: Region
; LOCATION: 33..56
; OTHER INFORMATION: /label= Tm
; OTHER INFORMATION: /note= "Transmembrane region of App"
;
; FEATURE:
; NAME/KEY: Region
; LOCATION: 1..32
; OTHER INFORMATION: /label= Ex
; OTHER INFORMATION: /note= "N-terminal extracellular part of App"
;
US-08-472-627-3

Query Match      81.8%; Score 45; DB 1; Length 59;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  2 GGVVIATVIV 11
Db   41 GGVVIATVIV 50

RESULT 14
US-08-388-463-3
; Sequence 3, Application US/08388463
; Patent No. 5786180
; GENERAL INFORMATION:
; APPLICANT: Konig, Gerhard
; APPLICANT: Graham, Paul
; TITLE OF INVENTION: Monoclonal Antibody Specific for BA4
; TITLE OF INVENTION: Peptide
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Allegrretti & Witcoff, Ltd.
; STREET: 10 South Wacker Drive Suite 3000
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/388.463
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McDonnell, John J
; REGISTRATION NUMBER: 26,949
; REFERENCE/DOCKET NUMBER: 95,216
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-715-1000
; TELEFAX: 312-715-1234
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 59 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
```



```

; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Cleavage-site
; LOCATION: 4..5
; OTHER INFORMATION: /label= Beta
; OTHER INFORMATION: /note= "Beta cleavage site in App"
; FEATURE:
; NAME/KEY: Cleavage-site
; LOCATION: 20..21
; OTHER INFORMATION: /label= Alpha
; OTHER INFORMATION: /note= "Alpha cleavage site in APP, residues 16/17
; OTHER INFORMATION: of BA4."
; FEATURE:
; NAME/KEY: Cleavage-site
; LOCATION: 46..47
; OTHER INFORMATION: /label= Gamma
; OTHER INFORMATION: /note= "Gamma cleavage site in App"
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 5..47
; OTHER INFORMATION: /label= BA4
; OTHER INFORMATION: /note= "BA4 peptide"
; FEATURE:
; NAME/KEY: Region
; LOCATION: 33..56
; OTHER INFORMATION: /label= Tm
; OTHER INFORMATION: /note= "Transmembrane region of App"
; FEATURE:
; NAME/KEY: Region
; LOCATION: 1..32
; OTHER INFORMATION: /label= Ex
; OTHER INFORMATION: /note= "N-terminal extracellular part of App"
US-08-388-463-3

```

```

Query Match      81.8%; Score 45; DB 1; Length 59;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      2 GGVVIATVIV 11
Db      41 GGVVIATVIV 50

```

```

RESULT 15
US-08-462-859A-3
; Sequence 3, Application US/08462859A
; Patent No. 5652092
; GENERAL INFORMATION:
; APPLICANT: Jacobsen, J. S.
; APPLICANT: Vitek, M. P.
; TITLE OF INVENTION: No. 5652092el Amyloid Precursor and Method of
; TITLE OF INVENTION: Using Same to Access Agents Which Down-Regulate Formation
; TITLE OF INVENTION: of B-Amyloid Peptide
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Cyanamid Company
; STREET: One Cyanamid Plaza
; CITY: Wayne
; STATE: New Jersey
; COUNTRY: United States
; ZIP: 07470-8426
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462.859A
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Barnhard, Elizabeth M.

```

```

; REGISTRATION NUMBER: 31,088
; REFERENCE/DOCKET NUMBER: 31,844-04
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201)831-3246
; TELEFAX: (201)831-3305
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 63 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-462-859A-3

```

```

Query Match      81.8%; Score 45; DB 1; Length 63;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      2 GGVVIATVIV 11
Db      46 GGVVIATVIV 55

```

```

Search completed: September 6, 2001, 16:39:36
Job time: 125 sec

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:43:27 ; Search time 225.25 Seconds
(without alignments)
3.230 Million cell updates/sec

Title: US-09-603-713-7
Perfect score: 55
Sequence: 1 KGGWVATVIVK 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_0601.*
1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	55	100.0	12	22	Peptide derived fr
2	55	100.0	12	22	Peptide from gamma
3	45	81.8	23	19	Fluorogenic protea
4	45	81.8	24	15	Membrane-spanning
5	45	81.8	24	21	Transgenic APP pro
6	45	81.8	44	19	Human ALZASP2. Ho
7	45	81.8	45	18	Amyloid precursor
8	45	81.8	45	18	Amyloid precursor
9	45	81.8	45	19	APP-REP 751 [BAP d
10	45	81.8	45	19	Deletion beta-amy
11	45	81.8	47	20	Synthetic amyloid

12	45	81.8	49	14	AAR35087	Sequence encoded b
13	45	81.8	52	16	AAR64166	Variant beta amylo
14	45	81.8	52	20	AAW81476	Synthetic amyloid
15	45	81.8	53	15	AAR55695	Sequence of uniden
16	45	81.8	53	15	AAR55697	Sequence of uniden
17	45	81.8	53	21	AAW87944	Mammalian amyloid
18	45	81.8	54	21	AAW32126	Amyloid-beta precu
19	45	81.8	55	22	AAW11482	Human APP peptide
20	45	81.8	59	17	AAW05375	Amyloid precursor
21	45	81.8	59	19	AAW70863	Beta-amyloid precu
22	45	81.8	60	21	AAW69701	Beta-amyloid precu
23	45	81.8	63	18	AAW26511	Amyloid precursor
24	45	81.8	63	18	AAW26391	Amyloid precursor
25	45	81.8	63	19	AAW44747	APP-REP 751 [BAP pe
26	45	81.8	63	19	AAW44746	APP-REP 751 [BAP E
27	45	81.8	63	19	AAW42975	Beta-amyloid pepti
28	45	81.8	63	19	AAW42976	Beta-amyloid pepti
29	45	81.8	79	19	AAW53981	Human ALZASP. Hom
30	45	81.8	97	9	AAW81517	Deduced sequence o
31	45	81.8	97	9	AAW81512	Lambda SM2W3 encod
32	45	81.8	97	14	AAW37865	Beta-amyloid prote
33	45	81.8	99	13	AAW20329	Sequence of A99 (b
34	45	81.8	99	16	AAW64167	Variant beta amylo
35	45	81.8	99	16	AAW74694	Beta-amyloid precu
36	45	81.8	99	16	AAW74695	Beta-amyloid precu
37	45	81.8	99	20	AAW08606	Human beta-amyloid
38	45	81.8	99	22	AAW11483	Human APP peptide
39	45	81.8	100	12	AAW10024	Beta-amyloid-relat
40	45	81.8	100	14	AAW37866	Full-length beta-a
41	45	81.8	100	21	AAW13015	Human amyloid prec
42	45	81.8	100	21	AAW51923	Transgenic APP pro
43	45	81.8	103	16	AAW74697	Beta-amyloid precu
44	45	81.8	103	16	AAW74698	Beta-amyloid precu
45	45	81.8	103	19	AAW51317	Natural beta-amylo

ALIGNMENTS

RESULT 1
AAB66577
ID AAB66577 standard; Peptide; 12 AA.
XX AAB66577;
AC AAB66577;
DT 12-APR-2001 (first entry)
DE Peptide derived from APP gamma-secretase site.
XX Memapsin 2; nootropic; neuroprotective; amyloid precursor protein;
KW APP; memapsin 2 inhibitor; Alzheimer's disease.
XX Unidentified.
XX WO200100665-A2.
XX 04-JAN-2001.
XX 27-JUN-2000; 2000WO-US17742.
XX 28-JUN-1999; 99US-0141363.
PR 30-NOV-1999; 99US-0168060.
PR 25-JAN-2000; 2000US-0177836.
PR 27-JAN-2000; 2000US-0178368.
PR 08-JUN-2000; 2000US-0210292.
XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.
PA (UNII) UNIV ILLINOIS FOUND.
XX Tang JUN, Hong L, Ghosh AK;
DR WPI; 2001-137933/14.

Fri Sep 7 10:58:46 2001

PT Novel memapsin 2 inhibitors which bind to active site of memapsin 2
 PT having 2 catalytic aspartic residues and substrate binding cleft, used
 PT to treat Alzheimer's disease by blocking amyloid precursor protein
 XX cleavage
 XX
 PS Disclosure; Page 11; 86pp; English.
 XX
 CC The present sequence is given in a specification relating to an inhibitor
 CC of catalytically active memapsin 2. The inhibitor binds to the memapsin 2
 CC active site, which is defined by the presence of two catalytic aspartic
 CC residues and a substrate binding cleft. The inhibitor is useful for
 CC the treatment and diagnosis of Alzheimer's disease. It is useful in
 CC screens for individuals with a genetic predisposition to Alzheimer's
 CC disease. The inhibitor is useful as a reagent for specifically binding to
 CC memapsin 2 or memapsin 2 analogues and for aiding in memapsin 2
 CC isolation, purification and characterisation.
 XX
 XX Sequence 12 AA;

Query Match 100.0%; Score 55; DB 22; Length 12;
 Best Local Similarity 100.0%; Pred. No. 0.00053;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KGGVVIATVIVK 12
 |||||
 Db 1 kgvviatvirk 12

RESULT 2
 AAB61339
 ID AAB61339 standard; peptide; 12 AA.
 XX
 AC AAB61339;
 XX
 DT 02-APR-2001 (first entry)
 XX
 DE Peptide from gamma-secretase site of beta-amyloid precursor protein.
 XX
 KW Memapsin 2; catalyst; Alzheimer's.
 XX
 OS Unidentified.
 XX
 PN WO200100663-A2.
 XX
 PD 04-JAN-2001.
 XX
 PF 27-JUN-2000; 2000WO-US17661.
 XX
 PR 28-JUN-1999; 99US-0141363.
 PR 30-NOV-1999; 99US-0168060.
 PR 25-JAN-2000; 2000US-0177836.
 PR 27-JAN-2000; 2000US-0178368.
 PR 08-JUN-2000; 2000US-0210292.
 XX
 XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.
 XX
 XX Tang JJN, Lin X, Koelsch G;
 XX WPI; 2001-102885/11.
 XX
 XX Purified recombinant catalytically active memapsin 2, used to screen
 PT inhibitors of it, which are used to treat and prevent Alzheimer's
 PT disease -
 XX
 XX Disclosure; Page 11; 86pp; English.
 PS
 CC The present invention relates to a purified recombinant
 CC catalytically active memapsin 2. The invention may be used for
 CC isolating inhibitors which are used to treat or prevent
 CC Alzheimer's disease. The invention may also be used to screen
 CC for individuals more genetically prone to develop Alzheimer's
 CC disease.

XX Sequence 12 AA;
 SQ
 Query Match 100.0%; Score 55; DB 22; Length 12;
 Best Local Similarity 100.0%; Pred. No. 0.00053;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 KGGVVIATVIVK 12
 |||||
 Db 1 kgvviatvirk 12
 RESULT 3
 AAW82191
 ID AAW82191 standard; peptide; 23 AA.
 XX
 AC AAW82191;
 XX
 DT 18-FEB-1999 (first entry)
 XX
 DE Fluorogenic protease indicator APP[709-710] peptide.
 XX
 KW Protease activity; fluorophore; detection; fluorogenic; cellular uptake;
 KW conformation change.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Modified-site 3 /label= Alb
 FT /note= "alpha-aminoisobutyric acid, labelled as
 FT amino acid B in the specification"
 FT Modified-site 4 /note= "epsilon-aminocaproic acid, labelled as
 FT amino acid J in the specification"
 FT Modified-site 18 /note= "epsilon-aminocaproic acid, labelled as
 FT amino acid J in the specification"
 XX
 PN WO9837226-A1.
 XX
 PD 27-AUG-1998.
 XX
 PF 20-FEB-1998; 98WO-US03000.
 XX
 PR 20-FEB-1997; 97US-0802981.
 XX
 XX (ONCO-) ONCOIMMUNIN INC.
 XX
 XX Komoriya A, Packard BS;
 XX WPI; 1998-467579/40.
 XX
 XX New fluorogenic compositions - containing 2 fluorophores separated
 PT by a peptide comprising a protease binding site, used for detecting
 PT protease activity in samples.
 XX
 XX Disclosure; Page 29; 90pp; English.
 XX
 XX AAW82023-W82240 are peptides used in the construction of a fluorogenic
 CC composition which is used for the detection of protease activity in
 CC biological samples. The products can be used for the detection of
 CC conformation changes in nucleic acids, oligosaccharides,
 CC polysaccharides, proteins, peptides, lipids, phospholipids, glycolipids,
 CC glycoproteins, steroids or polymers. In addition, attachment of a
 CC hydrophobic group to a molecule can be used to enhance uptake by cells.
 CC The composition is composed of P = peptide comprising a protease binding
 CC site for the protease, F1, F2 peptides = fluorophores where F1 is
 CC attached to the amino terminal amino acid and F2 is attached to the
 CC carboxyl terminal amino acid and S1, S2 peptides = when present, are
 CC peptide spacers where S1, when present, is attached to the amino terminal
 CC acid, and S2, when present, is attached to the carboxyl terminal amino

CC acid.
XX Sequence 23 AA;
SQ

Query Match 81.8%; Score 45; DB 19; Length 23;
Best Local Similarity 100.0%; Pred. No. 0.072;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGVVIATVIV 11
DB 5 ggviatviv 14
|||||

RESULT 4
AAR58922
ID AAR58922 standard; peptide; 24 AA.
XX
AC AAR58922;
XX
DT 15-APR-1995 (first entry)
XX
DE Membrane-spanning segment of APP695 (residues 625-648).
XX
KW Amyloid precursor protein; APP695; beta amyloid; cytoplasmic tail;
KW Alzheimer's disease.
XX
OS Homo sapiens.
XX
PN W09419692-A.
XX
PD 01-SEP-1994.
XX
PF 17-FEB-1994; 94WO-US01712.
XX
PR 18-FEB-1993; 93US-0019208.
XX
PA (GEO) GEN HOSPITAL CORP.
XX
PI Nishimoto I;
XX
DR WPI; 1994-294486/36.
XX
PT Identifying cpds. useful for treating or preventing Alzheimer's
PT disease - by determining whether it interferes with the
PT association of the couplone portion of amyloid precursor protein
PT to G polypeptide
XX
PS Disclosure; Page 29; 71pp; English.
XX

Beta amyloid is synthesized as part of a larger protein referred to
as amyloid precursor protein (APP), which has a number of isoforms
in humans, including APP695 and APP770. The amino terminal of beta
amyloid is generated by cleavage of a peptide bond of APP which in
APP695 lies between Met596 and Asp597. APP forms a complex with Go,
a GTP-binding protein (or "G protein") in brain. Go is made of one
alpha subunit and one Beta-gamma subunit. Two isoforms of Go, known
as Go1 (or GoA) and Go2 (or GoB) have been identified; they have
slight AA differences in their alpha subunits. The cDNA sequence and
deduced AA sequence of the alpha subunits in each of Go1 and Go2 are
shown in AAQ69002/R58914 and AAQ69004/R58924 respectively. The
cytoplasmic APP695 sequence His657-Lys676 (AAR58913) possesses a
specific Go-activating function, and is necessary for complex
formation of this APP with Go. The cytoplasmic tail portion of
APP695 from residues 649-695 and the membrane-spanning
portion of APP695 from residues 639-648, and the entire
membrane-spanning segment from residues 625-648 are given in
AAR58917, AAR58918, and AAR58922 respectively. The sequence of all of
APP is given in AAQ69003/R58923.

Sequence 24 AA;
SQ

Query Match 81.8%; Score 45; DB 15; Length 24;
Best Local Similarity 100.0%; Pred. No. 0.075;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGVVIATVIV 11
DB 9 ggviatviv 18
|||||

RESULT 5
AAY51920
ID AAY51920 standard; peptide; 24 AA.
XX
AC AAY51920;
XX
DT 22-JUN-2000 (first entry)
XX
DE Transgenic APP protein fragment #1.
XX
KW APP; amyloid precursor protein; gamma-secretase; neuroprotective;
KW nontropic; transgenic; Alzheimer's disease; Down's syndrome.
XX
OS Homo sapiens.
XX
PN DE19856261-C1.
XX
PD 30-MAR-2000.
XX
PF 07-DEC-1998; 98DE-1056261.
XX
PR 07-DEC-1998; 98DE-1056261.
XX
PA (HMRI) HOECHST MARION ROUSSEL DEUT GMBH.
XX
PI Peraus G;
XX
DR WPI; 2000-258119/23.
XX
PT Detection of gamma-secretase by detection of A-beta peptide useful for
PT determining gamma-secretase activity and for identifying inhibitors -
XX
PS Claim 1.1a; Page 6; 16pp; German.
XX

This invention describes a novel method for the detection of human
gamma-secretase by detection of a partial protein formed by cleavage of
a fusion protein encoded by a transgene containing a first nucleotide
sequence which encodes a protein comprising the amino acid sequence (A)
and a second nucleotide sequence which encodes a signal peptide. The
products of the invention have neuroprotective and nontropic activity.
The method is used to detect activity of gamma-secretase. The transgene
and/or vectors are useful for the production of a transgenic cell or
C. elegans. Transgenic C. elegans is useful in a method for the
determination of gamma-secretase activity. The transgenic C. elegans is
also useful in a method to identify inhibitors of the gamma-secretase
activity. The methods and transgenes are useful in research of
Alzheimer's disease. Inhibitors of gamma-secretase are useful in
control/treatment of Alzheimer's and possibly Down's syndrome. This
sequence represents a fragment of transgenic amyloid precursor protein
(APP) which is described in the method of the invention.

Sequence 24 AA;
SQ

Query Match 81.8%; Score 45; DB 21; Length 24;
Best Local Similarity 100.0%; Pred. No. 0.075;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGVVIATVIV 11
DB 9 ggviatviv 18
|||||

RESULT 6

AAW53985

ID AAW53985 standard; Protein; 44 AA.

XX AC AAW53985;

XX DT 18-AUG-1998 (first entry)

XX DE Human ALZAsp2.

XX KW Dsas; DSAsp; alzas; Down's syndrome; diagnosis; therapy; human;

XX KW Alzheimer's disease.

XX OS Homo sapiens.

XX PN WO9807850-A2.

XX PD 26-FEB-1998.

XX PF 22-AUG-1997; 97WO-EP04599.

XX PR 22-AUG-1996; 96CA-2183901.

XX FA (BERG/) BERGMANN J E.

XX FA (PRED/) PREDDIE E R.

XX PI Bergmann JE, Preddie ER;

XX DR WPI; 1998-169155/15.

XX DR N-PSDB; AAV23755.

XX PT Nucleic acid molecules dsas, and alzas - used for detecting and
treating Down's syndrome and Alzheimer's disease
XX PS Claim 13; Fig 1M; 96pp; English.
XX CC This sequence is the ALZAsp2 encoded by the nucleic acid alzas. The dsas
and alzas DNA sequences are the nucleic acids of the invention. Reagents
specifically for DSAsp can be used for the diagnosis of Down's syndrome
in humans and especially in pregnant women. Molecules that inhibit the
activity of the promoters (PDS1, PDS2, and PDS4) for dsas can be
used for treating Down's syndrome. The reagent capable of detecting
alzas can be used for detecting Alzheimer's disease, especially in the
pre-symptomatic stage. Substances that inhibit the promoters for alzas
can be used in treating Alzheimer's disease.

SQ Sequence 44 AA;

Query Match 81.8%; Score 45; DB 19; Length 44;

Best Local Similarity 100.0%; Pred. No. 0.15;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGVVIATVIV 11

Db 3 ggviatviv 12

RESULT 7

AAW26512

ID AAW26512 standard; Peptide; 45 AA.

XX AC AAW26512;

XX DT 06-JAN-1998 (first entry)

XX DE Amyloid precursor protein fragment APP-REP 751 (BAP del11-28).

XX KW Amyloid precursor protein; APP; beta-amyloid protein; BAP;

XX KW substrate; mutein; secretase; Alzheimer's disease; human.

XX OS Chimeric Homo sapiens.

XX OS Chimeric synthetic.

XX FH

FH

FT Cleavage-site 7..8

FT FT /note= "secretase cleavage site"

FT FT 10..33

FT FT /label= BAP(del11-28)

FT FT /note= "truncated beta-amyloid protein"

FT FT 20..42

FT FT /label= Transmembrane

XX US5656477-A.

XX PN 12-AUG-1997.

XX PD 01-MAY-1992; 92US-0877675.

XX PF 20-SEP-1993; 93US-0123659.

XX PR 01-MAY-1992; 92US-0877675.

XX XX (AMCY) AMERICAN CYANAMID CO.

XX XX Jacobsen JS, Vitek MP;

XX PI WPI; 1997-414594/38.

XX DR Nucleic acid encoding amyloid precursor mutin(s) - comprising

XX PT reporter gene and coding sequence, for identifying compounds which

XX PT modify the activity of proteolytic enzymes which cleave APP

XX PS Disclosure; Fig 5A; 84pp; English.

XX CC This peptide sequence shows the region of amyloid precursor protein

XX CC (APP) that includes a truncated beta-amyloid protein (BAP) lacking

XX CC the native secretase cleavage/recognition site. In an attempt to

XX CC engineer an APP non-cleavable substrate for secretase, an

XX CC APP-reporter (APP-REP) protein that carries the BAP deletion has

XX CC been expressed in recombinant host cells. Deletion of these 18

XX CC amino acids, however, still resulted in the secretion of an

XX CC N-terminal APP-reporter fragment into the cytoplasm. Non-

XX CC cleavable APP substrates can be used to detect other putative

XX CC abnormal APP processing events. They can also be used to

XX CC investigate cellular post-translational modifications to APP in

XX CC order to determine the potential influence on normal secretase and

XX CC abnormal BAP 'clipping' activities.

XX SQ Sequence 45 AA;

Query Match 81.8%; Score 45; DB 18; Length 45;

Best Local Similarity 100.0%; Pred. No. 0.15;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGVVIATVIV 11

Db 28 ggviatviv 37

RESULT 8

AAW26392

ID AAW26392 standard; Peptide; 45 AA.

XX AC AAW26392;

XX DT 15-DEC-1997 (first entry)

XX DE Amyloid precursor protein fragment APP-REP 751 (BAP del11-28).

XX KW Amyloid precursor protein; APP; beta-amyloid protein; BAP;

XX KW substrate; mutein; secretase; Alzheimer's disease; human.

XX OS Chimeric Homo sapiens;

XX OS Chimeric synthetic.

XX FH

Key Location/Qualifiers

```

FT Cleavage-site 7..8 /note= "secretase cleavage site"
FT Peptide 10..33 /label= BAP(delta11-28)
FT Domain 20..42 /note= "truncated beta-amyloid protein"
FT /label= Transmembrane
XX
XX US5652092-A.
XX
XX 29-JUL-1997.
XX
XX 01-MAY-1992; 92US-0877675.
XX
XX 20-SEP-1993; 93US-0123659.
XX
XX 01-MAY-1992; 92US-0877675.
XX
XX 05-JUN-1995; 95US-0462859.
XX
XX (AMCY ) AMERICAN CYANAMID CO.
XX
XX Jacobsen JS, Vitek MP;
XX
XX WPI; 1997-392937/36.
XX
XX Screening for compounds which reduce beta-amyloid protein formation
XX - using cells which express a construct encoding a marker and an
XX amyloid precursor muten derived from APP isoforms
XX
XX Disclosure; Fig 5A; 84pp; English.
XX
XX This peptide sequence shows the region of amyloid precursor protein
XX (APP) that includes a truncated beta-amyloid protein (BAP) lacking
XX the native secretase cleavage/recognition site. In an attempt to
XX engineer an APP non-cleavable substrate for secretase, an
XX APP-reporter (APP-REP) protein that carries the BAP deletion has
XX been expressed in recombinant host cells. Deletion of these 18
XX amino acids, however, still resulted in the secretion of an
XX N-terminal APP-reporter fragment into the cytoplasm. Non-
XX cleavable APP substrates can be used to detect other putative
XX abnormal APP processing events. They can also be used to
XX investigate cellular post-translational modifications to APP in
XX order to determine the potential influence on normal secretase and
XX abnormal BAP 'clipping' activities.
XX
XX Sequence 45 AA;

Query Match 81.8%; Score 45; DB 18; Length 45;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGVVIATVIV 11
DB 28 ggviatviv 37

RESULT 9
AAW4748
ID AAW4748 standard; Protein; 45 AA.
XX
XX AAW4748;
XX
XX 01-JUN-1998 (first entry)
XX
XX APP-REP 751 [BAP delta(11-28)] peptide.
XX
XX Amyloid precursor protein; APP; APP 751 isoform; deletion; substrate P;
KW epitope; Met-enkephalin; detection; secretase; beta-amyloid protein; BAP;
KW Alzheimer's disease; cleavage.
XX
XX Homo sapiens.
XX Synthetic.
XX

```

```

FH Key Location/Qualifiers
FT Cleavage-site 7..8 /note= "putative secretase cleavage site"
FT Misc-difference 19..20 /note= "residues 11-28 of the wild type BAP sequence
FT are deleted from between these positions"
XX
XX US5693478-A.
XX
XX 02-DEC-1997.
XX
XX 05-JUN-1995; 95US-0464247.
XX
XX 20-SEP-1993; 93US-0123659.
XX
XX 01-MAY-1992; 92US-0877675.
XX
XX 05-JUN-1995; 95US-0464247.
XX
XX (AMCY ) AMERICAN CYANAMID CO.
XX
XX Jacobsen JS, Vitek MP;
XX
XX WPI; 1998-031744/03.
XX
XX Amyloid precursor muten reporter molecule assay containing antibody
XX recognised marker - used to study pathways associated with
XX Alzheimer's disease
XX
XX Disclosure; Fig 5A; 84pp; English.
XX
XX This sequence represent the beta-amyloid protein sequence from the
XX construct APP-REP751 [BAP delta(11-28)]. The mutant sequence contains
XX a deletion of the wild type BAP residues 11-28. This causes a
XX shortening of the BAP sequence. This may affect cleavage of the BAP by
XX the "secretase" dependent on whether the "secretase" recognises the
XX cleavage site by a positional effect or by sequence. The mutant sequence
XX can be used in a method to study secretase and beta-amyloid protein
XX (BAP)-generating pathways associated with Alzheimer's disease by
XX studying proteolytic cleavage of the reporter polypeptides (e.g.
XX AAW44744 and AAW44745).
XX
XX Sequence 45 AA;

Query Match 81.8%; Score 45; DB 19; Length 45;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGVVIATVIV 11
DB 28 ggviatviv 37

RESULT 10
AAW42977
ID AAW42977 standard; peptide; 45 AA.
XX
XX AAW42977;
XX
XX 01-MAY-1998 (first entry)
XX
XX Deletion beta-amyloid peptide (BAP) derived from APP-REP 751.
XX
XX Beta-amyloid peptide; BAP; extracellular BAP plaque;
KW cerebrovascular deposit; Alzheimers disease; Downs syndrome;
KW amyloid precursor protein; APP; secretase; BAP aggregation;
KW abnormal proteolytic cleavage.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX 20..43
XX Domain /note= "putative transmembrane domain"
XX

```

PN US5703209-A.
 PD 30-DEC-1997.
 PP 05-JUN-1995; 95US-0464248.
 PR 20-SEP-1993; 93US-0123659.
 PS 01-MAY-1992; 92US-0877675.
 PA (AMCY) AMERICAN CYANAMID CO.
 PI Jacobsen JS, Vitek MP;
 DR WPI; 1998-076482/07.
 PT Amyloid precursor protein fusion polypeptides - comprising APP
 fragment and marker, useful for research and drug screening
 PS Disclosure; Fig 5A; 84pp; English.
 CC The present sequence represents a beta-amyloid peptide (BAP), with
 CC a deletion amino acids 11-28 (numbered according to AAW42976). Abnormal
 CC accumulation of extracellular BAP in plaques and cerebrovascular
 CC deposits is characteristic in brains of individuals suffering from
 CC Alzheimer's disease and Down's syndrome. BAP is a poorly soluble,
 CC self-aggregating protein which is derived from a larger amyloid precursor
 CC protein (APP). APP is expressed as an integral membrane protein, and is
 CC cleaved by secretase, between BAP 16lys and 17leu. Cleavage at this site
 CC precludes amyloidogenesis and results in the release of the
 CC amino-terminal APP fragment. Three major isoforms of APP exist: APP-695,
 CC APP-751 and APP-770. These isoforms are derived by alternative splicing;
 CC APP-ARP 751 is a deletion construct of APP-751, which has a deletion of
 CC 276 amino acids to within 15 amino acids of the BAP domain. APP can be
 CC used as a substrate for studying abnormal proteolytic cleavage which
 CC results in the release of BAP, and also to screen for drugs that will
 CC inhibit such cleavage.
 XX
 SQ Sequence 45 AA;

Query Match 81.8%; Score 45; DB 19; Length 45;
 Best Local Similarity 100.0%; Pred. No. 0.15;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GGVVIATVIV 11
 Db 28 ggviatviv 37

RESULT 11
 AAW81475
 ID AAW81475 standard; peptide; 47 AA.
 AC AAW81475;
 XX
 XX 28-JAN-1999 (first entry)
 DT
 DE Synthetic amyloid beta (Abeta) peptide 10 (residues 1-47).
 XX
 XX Amyloid beta; Abeta; deoxygenated solvent; evaporative deposition;
 KW research; neurotoxicity; free-radical; glutamine synthetase.
 XX
 XX Synthetic.
 OS
 XX
 XX US5840838-A.
 PN
 XX
 XX 24-NOV-1998.
 PD
 XX
 XX 29-FEB-1996; 96US-0609090.
 PF
 XX
 XX 29-FEB-1996; 96US-0609090.
 PR
 XX
 XX (KENT) UNIV KENTUCKY RES FOUND.
 PA

XX Aksenov M, Butterfield DA, Carney JM, Hensley K;
 PI WPI; 1999-034120/03.
 XX
 XX Process for treating synthetic amyloid beta peptides - by organic
 PT solvent treatment, useful for studying neurotoxicity
 PT
 XX
 PS Claim 5; Columns 13-14; 14pp; English.
 XX
 CC Sequences AAW81466 to AAW81476 represent synthetic amyloid beta (Abeta)
 CC peptides. The invention provides a process for treating a synthetic
 CC Abeta peptide that comprises dissolving the peptide in a deoxygenated
 CC solvent selected from trifluoroethanol, hexafluorocyclohexane, dimethyl
 CC sulphoxide, morpholinopropanesulphonic acid, dimethylformamide and
 CC acetonitrile to a concentration of 0.01-10 mg/ml, incubating the
 CC solution at 20-65 deg. C for 0.5-4 hour, and removing the solvent by
 CC 'evaporative deposition', in 5-10 minutes. Synthetic amyloid beta
 CC peptides are useful as research tools for studying neurotoxicity.
 CC resulting from Abeta peptide-enhanced free-radical production. The
 CC treatment increases the activity of the synthetic Abeta peptides in tests
 CC to determine free-radical generating capacity and glutamine synthetase
 CC inactivation.
 XX
 SQ Sequence 47 AA;

Query Match 81.8%; Score 45; DB 20; Length 47;
 Best Local Similarity 100.0%; Pred. No. 0.16;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GGVVIATVIV 11
 Db 37 ggviatviv 46

RESULT 12
 AAR35087
 ID AAR35087 standard; Protein; 49 AA.
 AC AAR35087;
 XX
 XX 09-AUG-1993 (first entry)
 DT

DE Sequence encoded by exon 15, which encodes the C-terminal portion
 DE of amyloid beta-protein, amino acids 514 through 569 (exon 17 if
 DE numbered by the APP770 transcript).
 XX

KW Amyloid precursor protein; Alzheimer's disease; mutant; screening.
 XX
 XX Homo sapiens.
 OS

FH Key Location/Qualifiers
 FT Misc-difference 29 /note= "changed to Phe by G to T mutation"
 FT Misc-difference 26 /note= "carboxyl terminus of the longest
 FT beta-amyloid peptide sequence that has been
 FT reported".
 XX

PN WO9307296-A.
 XX
 XX 15-APR-1993.
 PD
 XX 05-OCT-1992; 92WO-US08470.
 PF
 XX 03-OCT-1991; 91US-0770581.
 PR
 XX (INDV) UNIV INDIANA FOUND.
 PA
 XX Benson MD, Farlow M, Ghetti B, Murrell J;
 PI WPI; 1993-134480/16.
 XX
 DR

DR N-PSDB; AAQ40217.
PT Method of screening for risk of Alzheimer's disease - by
PT detecting a guanine to thymine mutation at position 1924 of the
PT gene encoding amyloid precursor protein
XX
PS Disclosure; Fig 1; 21pp; English.
XX
CC A novel method for screening for risk of Alzheimer's disease
CC comprises assaying for a guanine to thymine point mutation at posn.
CC 1924 of a patient's gene encoding the amyloid precursor protein
CC (APP). The method of D. O. Wirak et al is used to generate a
CC fragment containing the mutant APP gene. Accordingly, an open
CC reading frame including the mutant APP is contained within a Bgl II-
CC Bam HI restriction fragment and generated by site directed
CC mutagenesis of APP cDNA sequences with a synthetic oligonucleotide
CC primer (AAQ40218).
XX
SQ Sequence 49 AA;

Query Match 81.8%; Score 45; DB 14; Length 49;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GGVVIATVIV 11
Db 20 ggvviatviv 29
|||||
20 ggvviatviv 29

RESULT 13
AAR64166
ID AAR64166 standard; peptide: 52 AA.
XX
AC AAR64166;
XX
DT 02-AUG-1995 (first entry)
XX
DE Variant beta amyloid protein plus transmembrane domain of APP.
XX
KW beta amyloid protein; mutant; variant; detection; amyloid deposition;
KW diagnosis; amyloidosis associated disease; Alzheimer's disease;
KW Down's syndrome; APP; amyloid precursor protein.
XX
OS Synthetic.
XX
FH Location/Qualifiers
FT Peptide 1..41
FT Peptide /note= "beta amyloid protein variant"
FT Peptide 42..52
FT Peptide /note= "part of the transmembrane domain of APP"
XX
PN W09428412-A.
XX
PD 08-DEC-1994.
XX
PF 27-MAY-1994; 94WO-US05809.
XX
PR 28-MAY-1993; 93US-0069010.
XX
PA (MIRI-) MIRIAM HOSPITAL.
XX
PI Majocha RE, Marotta CA;
XX
DR WPI; 1995-023013/03.
XX
PT Amyloid binding composition comprising labelled amyloid protein
PT and carrier - useful for in vivo imaging of amyloid deposits, for
PT diagnosing Alzheimer's disease and Down's Syndrome.
XX
PS Claim 5; Page 42-43; 58pp; English.
XX
CC AAR64166 shows the amino acid sequence of the beta amyloid protein

CC plus some of the amino acids of the amyloid precursor protein. The
CC protein binds amyloid and is useful for in vivo imaging of amyloid
CC deposits and hence diagnosis of an amyloidosis-associated disease, such
CC as Alzheimer's disease or Down's syndrome. AAR64165 shows the generic
CC sequence of the amyloid protein for generation of variants.
XX
SQ Sequence 52 AA;

Query Match 81.8%; Score 45; DB 16; Length 52;
Best Local Similarity 100.0%; Pred. No. 0.18;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GGVVIATVIV 11
Db 37 ggvviatviv 46
|||||
37 ggvviatviv 46

RESULT 14
AAW81476
ID AAW81476 standard; peptide: 52 AA.
XX
AC AAW81476;
XX
DT 28-JAN-1999 (first entry)
XX
DE Synthetic amyloid beta (Abeta) peptide 11 (residues 1-52).
XX
KW Amyloid beta; Abeta; deoxygenated solvent; evaporative deposition;
KW research; neurotoxicity; free-radical; glutamine synthetase.
XX
OS Synthetic.
XX
PN US5840838-A.
XX
PD 24-NOV-1998.
XX
PF 29-FEB-1996; 96US-0609090.
XX
PR 29-FEB-1996; 96US-0609090.
XX
PA (KENT) UNIV KENTUCKY RES FOUND.
XX
PI Aksenov M, Butterfield DA, Carney JM, Hensley K;
XX
DR WPI; 1999-034120/03.
XX
PT Process for treating synthetic amyloid beta peptides - by organic
PT solvent treatment, useful for studying neurotoxicity
XX
PS Claim 5; Columns 13-14; 14pp; English.
XX
CC Sequences AAW81466 to AAW81476 represent synthetic amyloid beta (Abeta)
CC peptides. The invention provides a process for treating a synthetic
CC Abeta peptide that comprises dissolving the peptide in a deoxygenated
CC solvent selected from trifluoroethanol, hexafluorocyclohexane, dimethyl
CC sulphoxide, morpholinopropanesulphonic acid, dimethylformamide and
CC acetonitrile to a concentration of 0.01-10 mg/ml, incubating the
CC solution at 20-65 deg. C for 0.5-4 hour, and removing the solvent by
CC 'evaporative deposition', in 5-10 minutes. Synthetic amyloid beta
CC peptides are useful as research tools for studying neurotoxicity
CC resulting from Abeta peptide -enhanced free-radical production. The
CC treatment increases the activity of the synthetic Abeta peptides in tests
CC to determine free-radical generating capacity and glutamine synthetase
CC inactivation.
XX
SQ Sequence 52 AA;

Query Match 81.8%; Score 45; DB 20; Length 52;
Best Local Similarity 100.0%; Pred. No. 0.18;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GGVIATVIV 11
Db 37 ggviatviv 46

RESULT 15
AAR55695
ID AAR55695 standard; Protein; 53 AA.
XX
AC AAR55695;
XX
DT 29-DEC-1994 (first entry)
XX
DE Sequence of unidentified protein sequence ID number 20.
XX
KW Amyloid precursor protein; APP; plaque; beta/A4; Alzheimers;
KW transgene; ss.
XX
OS Synthetic.
XX
PN WO9412627-A.
XX
PD 09-JUN-1994.
XX
PF 24-NOV-1993; 93WO-US11480.
XX
PR 25-NOV-1992; 92US-0988850.
PR 09-NOV-1993; 93US-0149222.
XX
PA (CEPH-) CEPHALON INC.
XX
PI Howland DS, Scott RW;
XX
DR WPI; 1994-200256/24.
XX

PT Transgenic animal model for Alzheimer's disease - contains
PT transgene encoding amyloid protein under transcriptional control
PT of synapsin gene promoter
XX
PS Example; Page 54; 94pp; English.
XX
CC There was no apparent reference to sequence ID numbers 20, 21 or 22
CC in the specification. The specification describes a transgenic
CC animal model for Alzheimer's disease. A transgenic animal
CC harbouring a transgene coding for an amyloid protein under the
CC control of a promoter is claimed. The amyloid protein can be
CC APP695, APP751 or APP770. The coding sequence may contain a
CC mutation, including the hereditary cerebral haemorrhage with
CC amyloidosis - Dutch type (HCHWA-DT) and familial Alzheimer's
CC disease (FAD). Perhaps SQ IDs 20-22 corresp. to wt, HCHWA-DT
CC and FAD?
XX
SQ Sequence 53 AA;

Query Match 81.8%; Score 45; DB 15; Length 53;
Best Local Similarity 100.0%; Pred. No. 0.18;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GGVIATVIV 11
Db 41 ggviatviv 50

Search completed: September 6, 2001, 16:43:28
Job time: 357 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:43:27 ; Search time 225.25 Seconds
(without alignments)
2.153 Million cell updates/sec

Title: US-09-603-713-6

Perfect score: 40

Sequence: 1 SVNMAEGD 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_0601.*
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	100.0	8	22 AAB66576	Peptide derived fr
2	40	100.0	8	22 AAB61338	peptide from proce
3	36	90.0	8	22 AAB66579	Synthetic peptide
4	36	90.0	8	22 AAB61341	Recombinant human
5	36	90.0	101	21 AAB53687	Human colon cancer
6	36	90.0	125	18 AAW11782	Early onset Alzhei
7	36	90.0	164	18 AAW11787	Early onset Alzhei
8	36	90.0	407	18 AAW28507	Partial AD3 sequen
9	36	90.0	429	19 AAW41429	PS1/429 protein.
10	36	90.0	430	17 AAW05761	Presenilin-1-2 D25
11	36	90.0	434	17 AAW05760	Presenilin-1-1 D25

12	36	90.0	463	17 AAW05734	Presenilin-1-2. H
13	36	90.0	463	18 AAW22948	Presenilin-1 VRSQ
14	36	90.0	463	18 AAW12376	Human S182 gene pr
15	36	90.0	463	18 AAW11840	Early onset Alzhei
16	36	90.0	463	19 AAY20854	Human presenilin 1
17	36	90.0	463	19 AAW23965	Human presenilin-1
18	36	90.0	463	19 AAW42375	Human presenilin 1
19	36	90.0	465	21 AAY51393	Human S182 protein
20	36	90.0	465	21 AAY51394	Murine S182 protei
21	36	90.0	467	17 AAW05750	Presenilin-1-1 C26
22	36	90.0	467	17 AAW05751	Presenilin-1-1 P26
23	36	90.0	467	17 AAW05752	Presenilin-1-1 P26
24	36	90.0	467	17 AAW05753	Presenilin-1-1 E28
25	36	90.0	467	17 AAW05754	Presenilin-1-1 A28
26	36	90.0	467	17 AAW05755	Presenilin-1-1 L28
27	36	90.0	467	17 AAW05757	Presenilin-1-1 G38
28	36	90.0	467	17 AAW05758	Presenilin-1-1 L39
29	36	90.0	467	17 AAW05759	Presenilin-1-1 C41
30	36	90.0	467	17 AAW05733	Presenilin-1-1. H
31	36	90.0	467	17 AAW05735	Murine presenilin.
32	36	90.0	467	17 AAW05736	Presenilin-1-1 A79
33	36	90.0	467	17 AAW05737	Presenilin-1-1 V96
34	36	90.0	467	17 AAW05738	Presenilin-1-1 Y11
35	36	90.0	467	17 AAW05739	Presenilin-1-1 Y11
36	36	90.0	467	17 AAW05740	Presenilin-1-1 M13
37	36	90.0	467	17 AAW05741	Presenilin-1-1 L14
38	36	90.0	467	17 AAW05742	Presenilin-1-1 M14
39	36	90.0	467	17 AAW05743	Presenilin-1-1 H16
40	36	90.0	467	17 AAW05744	Presenilin-1-1 L17
41	36	90.0	467	17 AAW05745	Presenilin-1-1 G20
42	36	90.0	467	17 AAW05746	Presenilin-1-1 I21
43	36	90.0	467	17 AAW05747	Presenilin-1-1 I23
44	36	90.0	467	17 AAW05748	Presenilin-1-1 A24
45	36	90.0	467	17 AAW05749	Presenilin-1-1 A26

ALIGNMENTS

RESULT 1
AAB66576
ID AAB66576 standard; Peptide; 8 AA.
XX
AC AAB66576;
XX
DT 12-APR-2001 (first entry)
XX
DE Peptide derived from presenilin 1 processing site.
XX
KW Presenilin 1; memapsin 2; nontropic; neuroprotective;
KW amyloid precursor protein; APP; memapsin 2 inhibitor;
KW Alzheimer's disease.
XX
OS Unidentified.
XX
PN WO200100665-A2.
XX
XX 04-JAN-2001.
XX
PF 27-JUN-2000; 2000WO-US17742.
XX
PR 28-JUN-1999; 99US-0141363.
PR 30-NOV-1999; 99US-0168060.
PR 25-JAN-2000; 2000US-0177836.
PR 27-JAN-2000; 2000US-0178368.
PR 08-JUN-2000; 2000US-0210292.
XX
PA (OKLA-) OKLAHOMA MEDICAL RES FOUND.
PA (UNII) UNIV ILLINOIS FOUND.
XX
PI Tang JUN, Hong L, Ghosh AK;
XX
DR WPI; 2001-137933/14.

XX Novel memapsin 2 inhibitors which bind to active site of memapsin 2
 PT having 2 catalytic aspartic residues and substrate binding cleft, used
 PT to treat Alzheimer's disease by blocking amyloid precursor protein
 PT cleavage
 XX
 PS Disclosure; Page 11; 86pp; English.
 XX
 CC The present sequence is given in a specification relating to an inhibitor
 CC of catalytically active memapsin 2. The inhibitor binds to the memapsin 2
 CC active site, which is defined by the presence of two catalytic aspartic
 CC residues and a substrate binding cleft. The inhibitor is useful for
 CC the treatment and diagnosis of Alzheimer's disease. It is useful in
 CC screens for individuals with a genetic predisposition to Alzheimer's
 CC disease. The inhibitor is useful as a reagent for specifically binding to
 CC memapsin 2 or memapsin 2 analogues and for aiding in memapsin 2
 CC isolation, purification and characterisation.
 XX
 XX Sequence 8 AA;

Query Match 100.0%; Score 40; DB 22; Length 8;
 Best Local Similarity 100.0%; Pred. No. 3.4e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVNNAEGD 8
 Db 1 svnmaegd 8

RESULT 2
 AAB61338
 ID AAB61338 standard; peptide; 8 AA.

XX AAB61338;

XX AC (first entry)

XX DT Peptide from processing site of presenilin 1.

XX DE Memapsin 2; catalyst; Alzheimer's.

XX KW Unidentified.

XX OS WO200100663-A2.

XX PN 04-JAN-2001.

XX PD 27-JUN-2000; 2000WO-US17661.

XX PF 28-JUN-1999; 99US-0141363.

XX PR 30-NOV-1999; 99US-0168060.

XX PR 25-JAN-2000; 2000US-0177836.

XX PR 27-JAN-2000; 2000US-0178368.

XX PR 08-JUN-2000; 2000US-0210292.

XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.

XX Tang JJN, Lin X, Koelsch G;

XX WPI; 2001-102895/11.

XX Purified recombinant catalytically active memapsin 2, used to screen
 PT inhibitors of it, which are used to treat and prevent Alzheimer's
 PT disease -

XX PS Disclosure; Page 11; 86pp; English.

XX CC The present invention relates to a purified recombinant
 CC catalytically active memapsin 2. The invention may be used for
 CC isolating inhibitors which are used to treat or prevent
 CC Alzheimer's disease. The invention may also be used to screen
 CC for individuals more genetically prone to develop Alzheimer's

CC disease.
 XX Sequence 8 AA;

Query Match 100.0%; Score 40; DB 22; Length 8;
 Best Local Similarity 100.0%; Pred. No. 3.4e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVNNAEGD 8
 Db 1 svnmaegd 8

RESULT 3
 AAB66579
 ID AAB66579 standard; Peptide; 8 AA.

XX AAB66579;

XX DT 12-APR-2001 (first entry)

XX DE Synthetic peptide PSI-gamma.

XX KW Human; memapsin 2; nootropic; neuroprotective; amyloid precursor protein;
 KW APP; memapsin 2 inhibitor; Alzheimer's disease; PSI-gamma.

XX OS Homo sapiens.

XX OS Synthetic.

XX PN WO200100665-A2.

XX PD 04-JAN-2001.

XX PF 27-JUN-2000; 2000WO-US17742.

XX PR 28-JUN-1999; 99US-0141363.

XX PR 30-NOV-1999; 99US-0168060.

XX PR 25-JAN-2000; 2000US-0177836.

XX PR 27-JAN-2000; 2000US-0178368.

XX PR 08-JUN-2000; 2000US-0210292.

XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.

XX PA (UNII) UNIV ILLINOIS FOUND.

XX PI Tang JJN, Hong L, Ghosh AK;

XX WPI; 2001-137933/14.

XX Novel memapsin 2 inhibitors which bind to active site of memapsin 2
 PT having 2 catalytic aspartic residues and substrate binding cleft, used
 PT to treat Alzheimer's disease by blocking amyloid precursor protein
 PT cleavage -

XX Example 4; Page 33; 86pp; English.

XX The present sequence is given in a specification relating to an inhibitor
 CC of catalytically active memapsin 2. The inhibitor binds to the memapsin 2
 CC active site, which is defined by the presence of two catalytic aspartic
 CC residues and a substrate binding cleft. The inhibitor is useful for
 CC the treatment and diagnosis of Alzheimer's disease. It is useful in
 CC screens for individuals with a genetic predisposition to Alzheimer's
 CC disease. The inhibitor is useful as a reagent for specifically binding to
 CC memapsin 2 or memapsin 2 analogues and for aiding in memapsin 2
 CC isolation, purification and characterisation.

XX Sequence 8 AA;

Query Match 90.0%; Score 36; DB 22; Length 8;
 Best Local Similarity 100.0%; Pred. No. 3.4e+05;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VNMAEGD 8
|||||||

Db 2 vnmaegd 8

RESULT 4

AAB61341
ID AAB61341 standard; peptide: 8 AA.

XX AC AAB61341;

XX DT 02-APR-2001 (first entry)

XX DE Recombinant human memapsin 2 substrate.

XX KW Memapsin 2; catalyst; Alzheimer's.

XX OS Unidentified.

XX PN WO200100663-A2.

XX PD 04-JAN-2001.

XX PF 27-JUN-2000; 2000WO-US17661.

XX PR 28-JUN-1999; 99US-0141363.

XX PR 30-NOV-1999; 99US-0168060.

XX PR 25-JAN-2000; 2000US-0177836.

XX PR 27-JAN-2000; 2000US-0178368.

XX PR 08-JUN-2000; 2000US-0210292.

XX PA (OKLA-) OKLAHOMA MEDICAL RES FOUND.

XX PI Tang JUN, Lin X, Koelsch G;

XX DR WPI: 2001-102885/11.

XX PT Purified recombinant catalytically active memapsin 2, used to screen

XX PT inhibitors of it, which are used to treat and prevent Alzheimer's

XX PT disease -

XX PS Disclosure; Page 24; 86pp; English.

XX XX The present invention relates to a purified recombinant

XX CC catalytically active memapsin 2. The invention may be used for

XX CC isolating inhibitors which are used to treat or prevent

XX CC Alzheimer's disease. The invention may also be used to screen

XX CC for individuals more genetically prone to develop Alzheimer's

XX CC disease.

XX SQ Sequence 8 AA;

Query Match 90.0%; Score 36; DB 22; Length 8;

Best Local Similarity 100.0%; Pred. No. 3.4e+05;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VNMAEGD 8

|||||||

Db 2 vnmaegd 8

RESULT 5

AAB53687
ID AAB53687 standard; Protein; 101 AA.

XX AC AAB53687;

XX DT 09-MAR-2001 (first entry)

XX DE Human colon cancer antigen protein sequence SEQ ID NO:1227.

XX DE Human; colon cancer; colon cancer antigen; diagnosis; detection;

KW identification; cytostatic; cardioactive; neuroprotective; vulnery;

KW immunomodulatory; muscular; gynaecological; gastrointestinal;

KW nephrotropic; antinefactive; antibacterial; gene therapy; wound;

KW neural disorder; immune system disorder; muscular disorder;

KW reproductive disorder; gastrointestinal disorder; renal disorder;

KW infectious disease; cardiovascular disorder.

XX OS Homo sapiens.

XX PN WO200055351-A1.

XX PD 21-SEP-2000.

XX PF 08-MAR-2000; 2000WO-US05883.

XX PR 12-MAR-1999; 99US-0124270.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Rosen CA, Ruben SM;

XX DR WPI: 2000-587534/55.

XX DR N-PSDB; AAC98444.

XX PT Colon cancer associated gene sequences, referred to as colon cancer

XX PT antigens, useful for the treatment, prevention, and diagnosis of colon

XX PT disorders such as colon cancer -

XX PS Claim 11; Page 1815-1816; 2104pp; English.

XX CC AAC97991 to AAC98763 encode the human colon cancer associated proteins,

XX CC called human colon cancer antigens, given in AAB53334 to AAB54006. The

XX CC human colon cancer antigens can have cytostatic, cardioactive, muscular;

XX CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,

XX CC vulnery, nephrotropic, antinefactive and antibacterial activities, and

XX CC can be used in gene therapy. The colon cancer antigen polynucleotides,

XX CC proteins and antibodies to the proteins are useful for the prevention,

XX CC treatment and diagnosis of colon disorders, such as colon cancer. The

XX CC polynucleotides may be used in diagnostics and research, such as for

XX CC chromosome identification, and as hybridisation probes. The proteins

XX CC may also be used to prevent diseases such as neural disorders, immune

XX CC system disorders, muscular disorders, reproductive disorders,

XX CC gastrointestinal disorders, wounds, renal disorders, infectious

XX CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and

XX CC AAB54007 represent sequences used in the exemplification of the present

XX CC invention.

XX SQ Sequence 101 AA;

Query Match 90.0%; Score 36; DB 21; Length 101;

Best Local Similarity 100.0%; Pred. No. 1.2;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VNMAEGD 8

|||||||

Db 50 vnmaegd 56

RESULT 6

AAW11782

ID AAW11782 standard; Protein; 125 AA.

XX AC AAW11782;

XX DT 24-APR-1997 (first entry)

XX DE Early onset Alzheimer's disease gene product (T-cell lymphoma).

XX DE Early onset Alzheimer's disease; EOAD; diagnosis; therapy;

XX KW expressed sequence tag; EST; T-cell lymphoma.

XX OS Homo sapiens.

XX	Key	Location/Qualifiers	CC	(N.B. in-frame stop codons in the 6 reading frames of the cDNA clone are ignored in the translated polypeptide sequences given in the specification)
FT	Misc-difference 1	/note= "amino acid residue at position 1 is unidentified owing to degeneracy of the nucleotide sequence"	CC	
FT			CC	
FT			XX	
FT	Misc-difference 6	/note= "amino acid residue at position 6 is unidentified owing to degeneracy of the nucleotide sequence"	SQ	Sequence 125 AA;
FT				Query Match 90.0%; Score 36; DB 18; Length 125;
FT	Misc-difference 18	/note= "amino acid residue at position 18 is unidentified owing to degeneracy of the nucleotide sequence"		Best Local Similarity 100.0%; Pred. No. 1.6;
FT				Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
FT	Misc-difference 21	/note= "amino acid residue at position 21 is unidentified owing to degeneracy of the nucleotide sequence"	QY	2 VNMAEGD 8
FT			Db	
FT				30 vnmaegd 36
FT	Misc-difference 84	/note= "amino acid residue at position 84 is unidentified owing to degeneracy of the nucleotide sequence"	RESULT 7	
FT			AAW11787	
FT	Misc-difference 90	/note= "amino acid residue at position 90 is unidentified owing to degeneracy of the nucleotide sequence"	ID	AAW11787 standard; Protein; 164 AA.
FT			XX	
FT	Misc-difference 98	/note= "amino acid residue at position 98 is unidentified owing to degeneracy of the nucleotide sequence"	AC	AAW11787;
FT			XX	
FT	Misc-difference 113	/note= "amino acid residue at position 113 is unidentified owing to degeneracy of the nucleotide sequence"	DT	24-APR-1997 (first entry)
FT			XX	
FT	Misc-difference 120	/note= "amino acid residue at position 120 is unidentified owing to degeneracy of the nucleotide sequence"	DE	Early onset Alzheimer's disease gene product (activated T-cell).
FT			XX	
FT	Misc-difference 122	/note= "amino acid residue at position 122 is unidentified owing to degeneracy of the nucleotide sequence"	KW	Early onset Alzheimer's disease; EOAD; diagnosis: therapy;
FT			KW	expressed sequence tag; EST; activated T cell.
XX	WO9701573-A2.		XX	
XX			OS	Homo sapiens.
XX	16-JAN-1997.		XX	
XX	26-JUN-1996; 96WO-US11002.		Key	Location/Qualifiers
XX			FT	Misc-difference 144
XX	28-JUN-1995; 95US-0000590.		FT	/note= "amino acid residue at position 144 is unidentified owing to degeneracy of the nucleotide sequence"
XX	(SMIK) SMITHKLINE BEECHAM CORP.		FT	
XX	(SMIK) SMITHKLINE BEECHAM PLC.		FT	Misc-difference 151
XX	Karran EH;		FT	/note= "amino acid residue at position 151 is unidentified owing to degeneracy of the nucleotide sequence"
XX	WPI; 1997-100161/09.		FT	
XX	N-PSDB; AAT59472.		FT	Misc-difference 160
XX			FT	/note= "amino acid residue at position 160 is unidentified owing to degeneracy of the nucleotide sequence"
XX	New Early Onset Alzheimer's Disease gene - used to develop prods. for diagnosis, detection of pre-disposition to, or treatment of Alzheimer's disease		FT	
XX	Claim 3; Page 65-66; 97pp; English.		XX	WO9701573-A2.
XX			XX	
XX	A set of polypeptides (AAW11780-85) have sequences deduced from the 3 reading frames, of both strands, of a human T-cell lymphoma cDNA clone (AAT59472) that corresponds to an expressed sequence tag (EST) of the early onset Alzheimer's disease (EOAD) gene. EOAD gene products (see also AAW11768-79 and AAW11786-97) can be expressed in host cells and used to screen for agonists or antagonists useful in EOAD therapy, or to raise antibodies useful in the diagnosis of EOAD or predisposition to EOAD.		XX	16-JAN-1997.
CC			XX	26-JUN-1996; 96WO-US11002.
			XX	28-JUN-1995; 95US-0000590.
			XX	(SMIK) SMITHKLINE BEECHAM CORP.
			XX	(SMIK) SMITHKLINE BEECHAM PLC.
			XX	Karran EH;
			XX	WPI; 1997-100161/09.
			XX	N-PSDB; AAT59472.
			XX	New Early Onset Alzheimer's Disease gene - used to develop prods. for diagnosis, detection of pre-disposition to, or treatment of Alzheimer's disease
			XX	Claim 3; Page 69-70; 97pp; English.
			XX	A set of polypeptides (AAW11786-91) have sequences deduced from the 3 reading frames, of both strands, of a human activated T-cell cDNA clone (AAT59473) that corresponds to an expressed sequence tag (EST) of the early onset Alzheimer's disease (EOAD) gene. EOAD gene products (see also AAW11768-85 and AAW11792-97) can be expressed in host cells and used to screen for agonists or antagonists useful in EOAD therapy, or to raise antibodies useful in the diagnosis of EOAD or predisposition to EOAD.

CC EOAD therapy, or to raise antibodies useful in the diagnosis of
 CC EOAD or predisposition to EOAD.
 CC (N.B. in-frame stop codons in the 6 reading frames of the cDNA
 CC clone are ignored in the translated polypeptide sequences given in
 CC the specification)
 XX
 SQ Sequence 164 AA;

Query Match 90.0%; Score 36; DB 18; Length 164;
 Best Local Similarity 100.0%; Pred. No. 2.1;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 VNMAEGD 8
 |||||
 Db 5 vnmaegd 11

RESULT 8
 AAW28507
 ID AAW28507 standard; Protein: 407 AA.
 XX
 AC AAW28507;
 XX
 DT 07-DEC-1997 (first entry)
 XX
 DE Partial AD3 sequence.
 XX
 KW AD3; AD4/AD3LP; Alzheimer's disease; chromosome; missegregation;
 KW presenilin; inhibitor; AD; trisomy 21.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 86
 FT /label= mutation
 FT /note= "M -> L"
 FT
 FT Misc-difference 103
 FT /label= mutation
 FT /note= "H -> R"
 FT
 FT Misc-difference 186
 FT /label= mutation
 FT /note= "A -> E"
 FT
 FT Misc-difference 226
 FT /label= mutation
 FT /note= "L -> V"
 FT
 FT Misc-difference 350
 FT /label= mutation
 FT /note= "C -> Y"
 FT
 XX
 PN WO9707213-A2.
 XX
 XX 27-FEB-1997.
 XX
 XX 15-AUG-1996; 96WO-US13314.
 XX
 XX 16-AUG-1995; 95US-0002448.
 XX
 XX (HARD) HARVARD COLLEGE.
 XX
 XX Li J, Potter H;
 XX
 XX WPI; 1997-165297/15.
 DR N-PSDB; AAT87402.
 XX
 XX Identifying genes which cause chromosome missegregation - useful for
 PT identifying causes of and treatments for diseases, e.g. Alzheimer's
 PT disease, cancer and ageing
 XX
 PS Disclosure; Fig 1; 77pp; English.
 XX
 XX Identifying genes which cause improper chromosome segregation,
 CC screening for inhibitors of chromosome missegregation and processes

CC caused by genes encoding chromosome missegregation promoters
 CC was exemplified using Alzheimer's disease. The sequences
 CC given in AAT87401 to AAT87426 can be used in the above methods.
 CC The five mutations indicated in the Features Table cosegregate
 CC with early-onset familial Alzheimer's disease. It is predicted
 CC that these mutations result in increased levels of cells with
 CC trisomy 21 in carriers of the mutation compared with non-carriers.
 XX
 SQ Sequence 407 AA;

Query Match 90.0%; Score 36; DB 18; Length 407;
 Best Local Similarity 100.0%; Pred. No. 5.9;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 VNMAEGD 8
 |||||
 Db 236 vnmaeqd 242

RESULT 9
 AAW41429
 ID AAW41429 standard; Protein: 429 AA.
 XX
 AC AAW41429;
 XX
 DT 04-JUN-1998 (first entry)
 XX
 DE PS1/429 protein.
 XX
 KW Presenilin peptide; PS1/429; immunogen; immune response; PS1 gene;
 KW Alzheimer's disease; mitochondrial pathology; neurodegeneration;
 KW apoptosis.
 XX
 OS Homo sapiens.
 XX
 PN WO9746678-A1.
 XX
 PD 11-DEC-1997.
 XX
 PF 03-JUN-1997; 97WO-US09272.
 XX
 PR 18-JUL-1996; 96US-0683315.
 PR 06-JUN-1996; 96US-0659296.
 XX
 XX (FARB) BAYER CORP.
 XX
 XX Chisholm JC, Davis JN, Drache B;
 XX
 DR WPI; 1998-042186/04.
 DR N-PSDB; AAV17357.
 XX
 XX DNA encoding presenilin peptide PS1/429 and its analogues - useful
 PT for diagnosis and treatment of Alzheimer's disease
 XX
 PS Claim 24; Fig 1; 77pp; English.
 XX
 XX This sequence is the PS1/429 presenilin peptide (II) of the
 CC invention. Cells transformed with the DNA are used to produce recombinant
 CC (II) and analogues, useful e.g. as immunogens for generating an immune
 CC response against PS1/429. (II) is a new product of the PS1 gene,
 CC mutations in which cause Alzheimer's disease (AD). The nucleic acids are
 CC generally useful as probes for detection and quantification of PS1/429,
 CC particularly for diagnosis of AD, especially the target sequences that
 CC hybridise with probes are isolated for sequencing. Antibodies (Ab) can
 CC also be diagnosed at the protein level using Ab as immunoassay reagents.
 CC Ab can also be used to identify epitopes and for affinity purification of
 CC peptides. Antisense nucleic acid may also be used to regulate expression
 CC of the PS1/429 gene, and both nucleic acids and peptides are useful as
 CC size markers in electrophoresis, chromatography etc. The transgenic
 CC animals are used as models for AD, e.g. for testing drugs. Regulators of
 CC the PS1/429 gene or polypeptide can be used to treat e.g. AD or diseases
 CC involving mitochondrial pathology, apoptosis and neurodegeneration.

CC Typical regulators are antisense sequences, ribozymes, aptamers,
 CC synthetic or natural compounds. (If) may also be used to target other
 CC coding sequences to particular cellular locations.

XX
 SQ Sequence 429 AA;

Query Match 90.0%; Score 36; DB 19; Length 429;
 Best Local Similarity 100.0%; Pred. No. 6.3;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VNMAEGD 8
 |||||
 Db 258 vnmaegd 264

RESULT 10
 AAW05761
 ID AAW05761 standard; Protein; 430 AA.

XX AC AAW05761;

XX DT 25-JUL-1997 (first entry)

XX DE Presenilin-1-2 D253A delta254-286 mutation.

XX KW Presenilin-1; human; hPS1-1; hPS1-2; PS-2; integral membrane protein; AD;
 KW familial Alzheimer's disease; cerebral haemorrhage; schizophrenia;
 KW depression; antibody; gene expression modulator; therapy; muten.

XX OS Homo sapiens.

XX PH Key Location/Qualifiers
 FT Modified-site 253
 FT /label= D253A
 FT Misc-difference 253..254
 FT /note= "site of 33 residue deletion"

XX PN WO9634099-A2.

XX PD 31-OCT-1996.

XX PF 29-APR-1996; 96WO-CA00263.

XX PR 31-JUL-1995; 95US-0509359.

XX PR 28-APR-1995; 95US-0431048.

XX PR 28-JUN-1995; 95US-0496841.

XX PA (HSCR-) HSC RES & DEV LP.

XX PA (UTOR) UNIV TORONTO GOVERNING COUNCIL.

XX PI Fraser PE, Rommens JM, St George-Hyslop PH;

XX DR WPI; 1996-497631/49.

XX PT New presenilin genes - useful for diagnosis, therapy and drug

XX PT screening of familial Alzheimer's disease, cerebral disorders, etc.

XX PS Claim 3; Page -; 178pp; English.

XX CC This sequence represents mutated versions of the human presenilin-1-2
 CC protein (see AAW05734 for wild type sequence). AAW05734 represents a
 CC different wild type form of presenilin-1 that results from alternate
 CC splicing of the genomic DNA sequence. The presenilins are a family of
 CC highly conserved integral membrane proteins with a common structural
 CC motif, common alternate splicing patterns, and common mutational hot
 CC spot regions. Mutations in PS genes are implicated in familial
 CC Alzheimer's disease (AD) and possibly other diseases such as cerebral
 CC haemorrhage, schizophrenia, depression etc., so detection of mutations in
 CC the DNA encoding the wild type sequences can be used for diagnosis of
 CC these diseases. The wild type proteins, or vectors that express them or
 CC containing antisense sequences, antibodies selective for these mutant
 CC forms of the proteins and modulators of PS gene expression are

CC potentially useful for treatment of AD etc. Transgenic animals are
 CC useful as models for drug screening. The antibodies can also be used e.g.
 CC for affinity purification and in immunoassays.

XX SQ Sequence 430 AA;

Query Match 90.0%; Score 36; DB 17; Length 430;
 Best Local Similarity 100.0%; Pred. No. 6.3;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VNMAEGD 8
 |||||
 Db 259 vnmaegd 265

RESULT 11

AAW05760

ID AAW05760 standard; Protein; 434 AA.

XX AC AAW05760;

XX DT 23-JUL-1997 (first entry)

XX DE Presenilin-1-1 D257A delta258-290 mutation.

XX KW Presenilin-1; human; hPS1-1; hPS1-2; PS-2; integral membrane protein; AD;
 KW familial Alzheimer's disease; cerebral haemorrhage; schizophrenia;
 KW depression; antibody; gene expression modulator; therapy; muten.

XX OS Homo sapiens.

XX PH Key Location/Qualifiers
 FT Modified-site 257
 FT /label= D257A

FT Misc-difference 257..258
 FT /note= "site of 33 residue deletion"

XX PN WO9634099-A2.

XX PD 31-OCT-1996.

XX PF 29-APR-1996; 96WO-CA00263.

XX PR 31-JUL-1995; 95US-0509359.

XX PR 28-APR-1995; 95US-0431048.

XX PR 28-JUN-1995; 95US-0496841.

XX PA (HSCR-) HSC RES & DEV LP.

XX PA (UTOR) UNIV TORONTO GOVERNING COUNCIL.

XX PI Fraser PE, Rommens JM, St George-Hyslop PH;

XX DR WPI; 1996-497631/49.

XX PT New presenilin genes - useful for diagnosis, therapy and drug

XX PT screening of familial Alzheimer's disease, cerebral disorders, etc.

XX PS Claim 3; Page -; 178pp; English.

XX CC AAW05736-W05760 represent mutated versions of the human presenilin-1-1
 CC protein (see AAW05733 for wild type sequence). AAW05734 represents a
 CC different wild type form of presenilin-1 that results from alternate
 CC splicing of the genomic DNA sequence. The presenilins are a family of
 CC highly conserved integral membrane proteins with a common structural
 CC motif, common alternate splicing patterns, and common mutational hot
 CC spot regions. Mutations in PS genes are implicated in familial
 CC Alzheimer's disease (AD) and possibly other diseases such as cerebral
 CC haemorrhage, schizophrenia, depression etc., so detection of mutations in
 CC the DNA encoding the wild type sequences can be used for diagnosis of
 CC these diseases. The wild type proteins, or vectors that express them or
 CC containing antisense sequences, antibodies selective for these mutant
 CC forms of the proteins and modulators of PS gene expression are

CC potentially useful for treatment of AD etc. Transgenic animals are
 CC useful as models for drug screening. The antibodies can also be used e.g.
 CC for affinity purification and in immunoassays.

SQ Sequence 434 AA;

Query Match 90.0%; Score 36; DB 17; Length 434;
 Best Local Similarity 100.0%; Pred. No. 6.4;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 VNMAEGD 8
 Db 263 vnmaegd 269
 |||||

RESULT 12
 AAW05734
 ID AAW05734 standard; Protein; 463 AA.

XX AC AAW05734;

DT 23-JUL-1997 (first entry)

XX DE Presenilin-1-2.

XX KW Presenilin-1; human; hPS1-1; PS-2; integral membrane protein; AD;
 KW familial Alzheimer's disease; cerebral haemorrhage; schizophrenia;
 KW depression; antibody; gene expression modulator; therapy.

XX OS Homo sapiens.

XX PN W09634099-A2.

XX PD 31-OCT-1996.

XX PF 29-APR-1996; 96WO-CA00263.

XX PR 31-JUL-1995; 95US-0509359.

XX PR 28-APR-1995; 95US-0431048.

XX PR 28-JUN-1995; 95US-0496841.

XX PA (HSCR-) HSC RES & DEV LP.

XX PA (UTOR) UNIV TORONTO GOVERNING COUNCIL.

XX PI Fraser PE, Rommens JM, St George-Hyslop PH;

XX DR WPI; 1996-497631/49.

XX DR N-PSDB; AAT40029.

XX PT New presenilin genes - useful for diagnosis, therapy and drug

XX PT screening of familial Alzheimer's disease, cerebral disorders, etc.

XX PS Claim 2; Page 132-134; 178pp; English.

XX CC AAW05733 and AAW05734 represent the two different forms of wild type
 CC human presenilin-1 (PS-1). This form of presenilin-1 results from
 CC alternate splicing of the genomic DNA sequence. AAW05762 represents the
 CC coding sequence for wild type human PS-2. The presenilins are a family
 CC of highly conserved integral membrane proteins with a common structural
 CC motif, common alternate splicing patterns, and common mutational hot
 CC spot regions. Mutations in PS genes are implicated in familial
 CC Alzheimer's disease (AD) and possibly other diseases such as cerebral
 CC haemorrhage, schizophrenia, depression etc., so detection of mutations in
 CC the DNA encoding these sequences can be used for diagnosis of these
 CC diseases. These proteins, or vectors that express them or containing
 CC antisense sequences, antibodies selective for mutant forms of these
 CC proteins (such as AAW05736) and modulators of PS gene expression are
 CC potentially useful for treatment of AD etc. Transgenic animals are useful
 CC as models for drug screening. The antibodies can also be used e.g. for
 CC affinity purification and in immunoassays.

XX SQ Sequence 463 AA;

Query Match 90.0%; Score 36; DB 17; Length 463;
 Best Local Similarity 100.0%; Pred. No. 6.9;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 VNMAEGD 8
 Db 292 vnmaegd 298
 |||||

RESULT 13
 AAW22948
 ID AAW22948 standard; Protein; 463 AA.

XX AC AAW22948;

DT 19-MAR-1993 (first entry)

XX DE Presenilin-1 VRSQ variant.

XX KW Identification; determination; neurological disease susceptibility;
 KW detection; alternative splice site; polyadenylated mRNA transcript;
 KW familial Alzheimer's disease; FAD; presenilin 1; VRSQ variant.

XX OS Homo sapiens.

XX PN EP791660-A1.

XX PD 27-AUG-1997.

XX PF 14-FEB-1997; 97EP-0300988.

XX PR 22-FEB-1996; 96US-0012077.

XX PA (SMIK) SMITHKLINE BEECHAM CORP.
 PA (UYSF-) UNIV SOUTH FLORIDA DEPT PSYCHIATRY.
 PA (UNIW) UNIV WASHINGTON.

XX PI Barton A, Goate A, Hardy J;

XX DR WPI; 1997-418049/39.

XX DR N-PSDB; AAT75576.

XX PT Diagnosis of, or susceptibility to neurological disease -
 PT specifically Alzheimer's disease, by detecting aberrant splicing in
 PT mRNA

XX PS Example 1; Pages 9-11; 21pp; English.

XX CC Identifying susceptibility to a neurological disease, comprises
 CC detecting an alternative splice site in a polyadenylated mRNA
 CC transcript in a sample of genetic material, where the alternative
 CC splice site encodes AAW22944, or detecting AAW22944 in the protein
 CC encoded by the mRNA. Tests on 3 early onset familial Alzheimer's
 CC disease (FAD) patients, 6 late onset sporadic Alzheimer's disease
 CC (AD) patients and 4 neurologically normal subjects, indicated that
 CC mRNA transcripts of the presenilin 1 gene in samples from various
 CC brain regions occur in 2 forms, PS-1-long (containing a VRSQ
 CC motif) and PS-1-short (lacking the VRSQ motif, i.e. the protein
 CC denoted by the present sequence), and that the PS-1-long levels
 CC in hippocampus and frontal cortex samples are significantly lower
 CC in FAD patients than in AD and normal subjects.

XX SQ Sequence 463 AA;

Query Match 90.0%; Score 36; DB 18; Length 463;
 Best Local Similarity 100.0%; Pred. No. 6.9;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 VNMAEGD 8
 |||||

Db 292 vnmaegd 298

RESULT 14

AAW12376
ID AAW12376 standard; Protein; 463 AA.

XX AC AAW12376;

XX DT 17-JUN-1997 (first entry)

XX DE Human S182 gene product.

XX KW S182 gene; familial Alzheimer's disease; diagnosis;

XX KW transgenic animal.

XX OS Homo sapiens.

XX PN W09703999-Al.

XX PD 06-FEB-1997.

XX PF 26-JUN-1996; 96WO-US11065.

XX PR 02-AUG-1995; 95US-0001800.

XX PR 18-JUL-1995; 95US-0001500.

XX XX (UYSF-) UNIV SOUTH FLORIDA.

XX PA (UNIW) UNIV WASHINGTON SCHOOL MED.

XX PI Goate AM, Hardy JA;

XX DR WPI; 1997-132571/12.

XX DR N-PSDB; AAT63207.

XX PT New mutants of the S182 gene associated with familial Alzheimer's
PT disease - and related protein and transgenic animals, useful as
PT models for screening and assessing potential drugs

XX PS Disclosure; Fig 1A-D; 26pp; English.

XX CC A polypeptide (AAW12376) is the product of an S182 gene cDNA clone
CC (AAT63207) isolated from a human brain library. Several mutations in
CC the S182 gene have been found in families with members affected by
CC early onset Alzheimer's disease (AD): in 2 families Met to Val at
CC position 135; in 3 families Met to Val at 142; in 1 family Pro to
CC Ser at 263; in 4 families Glu to Ala at 276; and in 1 family Glu to
CC Gly at 280. Detection of the mutations is used to diagnose AD, or
CC a predisposition to it. Transgenic animals can be produced that
CC are useful as models for screening and assessing potential drugs.

XX SQ Sequence 463 AA;

Query Match

Best Local Similarity 90.0%; Score 36; DB 18; Length 463;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VNMAEGD 8

Db 292 vnmaegd 298

RESULT 15

AAW11840

ID AAW11840 standard; Protein; 463 AA.

XX AC AAW11840;

XX DT 07-MAY-1997 (first entry)

XX DE Early onset Alzheimer's disease (EOAD) splice variant polypeptide.

XX XX

KW Early onset Alzheimer's disease; EOAD; neurodegenerative disease;
KW diagnosis; therapy; inhibitor; antagonist; antibody.

XX OS Homo sapiens.

XX PN W09703086-Al.

XX PD 30-JAN-1997.

XX PF 26-JUN-1996; 96WO-US11064.

XX PR 18-JUL-1995; 95US-0001501.

XX PR 13-JUL-1995; 95US-0001142.

XX PA (UYSF-) UNIV SOUTH FLORIDA.

XX PI Hardy JA;

XX DR WPI; 1997-118980/11.

XX DR N-PSDB; AAT59536.

XX PT Early onset Alzheimer's disease gene - useful for diagnosing a
PT pre-disposition to Alzheimer's disease

XX PS Disclosure; Fig 2; 44pp; English.

XX CC A 463-amino acid polypeptide (AAW11840) is the product of a full-
CC length cDNA (AAT59536) of an early Onset Alzheimer's disease (EOAD)
CC splice variant gene. A 467-amino acid polypeptide (AAW11839) is the
CC product of a full-length cDNA (AAT59535) of the EOAD gene. The 2
CC polypeptides can be produced in transformed host cells and used to
CC raise antibodies, or to identify antagonist/inhibitor cpds. useful
CC in the treatment of Alzheimer's disease, esp. EOAD.

XX SQ Sequence 463 AA;

Query Match

Best Local Similarity 90.0%; Score 36; DB 18; Length 463;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VNMAEGD 8

Db 292 vnmaegd 298

Search completed: September 6, 2001, 16:43:27
Job time: 356 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:39:35 ; Search time 113.12 seconds
(without alignments)
1.456 Million cell updates/sec

Title: US-09-603-713-6
Perfect score: 40
Sequence: 1 SYNMAEGD 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	90.0	407	2	US-08-875-972-4
2	36	90.0	467	2	US-08-788-231A-15
3	36	90.0	463	2	US-08-670-479-18
4	36	90.0	463	3	US-08-670-964-4
5	36	90.0	463	3	US-08-888-077A-4
6	36	90.0	465	3	US-08-788-231A-17
7	36	90.0	467	2	US-08-967-101-2
8	36	90.0	467	2	US-08-967-101-4
9	36	90.0	467	2	US-08-967-101-134
10	36	90.0	467	2	US-08-592-541-2
11	36	90.0	467	2	US-08-592-541-4
12	36	90.0	467	2	US-08-592-541-134
13	36	90.0	467	3	US-08-923-454A-10
14	36	90.0	467	3	US-08-670-964-2
15	36	90.0	467	3	US-08-888-077A-2
16	36	90.0	467	3	US-08-888-077A-17
17	36	90.0	467	3	US-09-124-698-2
18	36	90.0	467	3	US-09-124-698-4
19	36	90.0	467	3	US-09-124-698-134
20	36	90.0	467	4	US-09-127-480-2
21	36	90.0	467	4	US-09-127-480-4
22	36	90.0	467	4	US-09-127-480-134
23	36	90.0	467	4	US-08-496-841C-2
24	36	90.0	467	4	US-08-496-841C-4
25	36	90.0	467	4	US-08-496-841C-134
26	36	90.0	467	4	US-08-496-841C-136
27	36	90.0	467	4	US-08-706-344C-2

28	36	90.0	467	4	US-08-706-344C-4	Sequence 4, Appl
29	36	90.0	467	4	US-08-706-344C-28	Sequence 28, Appl
30	36	90.0	467	4	US-08-706-344C-30	Sequence 30, Appl
31	36	90.0	467	4	US-08-706-344C-32	Sequence 32, Appl
32	31	77.5	679	1	US-08-214-583-2	Sequence 2, Appl
33	28	70.0	250	1	US-08-378-761A-78	Sequence 78, Appl
34	28	70.0	250	1	US-08-485-286-78	Sequence 78, Appl
35	28	70.0	263	1	US-07-901-707-7	Sequence 7, Appl
36	28	70.0	263	1	US-07-988-430-7	Sequence 7, Appl
37	28	70.0	263	1	US-08-425-336-7	Sequence 7, Appl
38	28	70.0	263	1	US-08-488-113B-7	Sequence 7, Appl
39	28	70.0	263	1	US-08-477-484B-7	Sequence 7, Appl
40	28	70.0	263	2	US-08-646-360-7	Sequence 7, Appl
41	28	70.0	263	4	US-08-839-765-7	Sequence 7, Appl
42	28	70.0	263	4	US-09-136-389-7	Sequence 7, Appl
43	28	70.0	263	5	PCT-US92-09487-7	Sequence 7, Appl
44	28	70.0	342	1	US-08-244-646-15	Sequence 15, Appl
45	28	70.0	342	1	US-08-592-936B-21	Sequence 21, Appl

ALIGNMENTS

RESULT 1
US-08-875-972-4
; Sequence 4, Application US/08875972
; Patent No. 5985564
; GENERAL INFORMATION:
; APPLICANT: Huntington Potter and Jinhue Li
; TITLE OF INVENTION: ASSAY FOR IDENTIFYING GENES CAUSING
; TITLE OF INVENTION: CHROMOSOME NON-DISJUNCTION
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173-4799
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/875,972
; FILING DATE: 08-AUG-97
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/002,448
; FILING DATE: 16-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan Esq., Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: H095-03PA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781) 861-6240
; TELEFAX: (781) 861-9540
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 407 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-875-972-4

Query Match 90.0%; Score 36; DB 2; Length 407;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VNMAGEGD 8
|||||

Db 236 VNMAEGD 242

RESULT 2

US-08-788-231A-15

; Sequence 15, Application US/08788231A

; Patent No. 6019974

; GENERAL INFORMATION:

; APPLICANT: L'Hernault, Steven W.

; TITLE OF INVENTION: SPE-4 RELATED PEPTIDES, ANTIBODIES AND

; TITLE OF INVENTION: METHODS

; NUMBER OF SEQUENCES: 20

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.

; STREET: 5370 Manhattan Circle, Suite 201

; CITY: Boulder

; STATE: Colorado

; COUNTRY: US

; ZIP: 80303

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/788,231A

; FILING DATE: 24-JAN-1997

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/010,672

; FILING DATE: 26-JAN-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Ferber, Donna M.

; REGISTRATION NUMBER: 33,878

; REFERENCE/DOCKET NUMBER: 60-95

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (303) 499-8080

; TELEFAX: (303) 499-8089

; INFORMATION FOR SEQ ID NO: 15:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 462 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: not relevant

; MOLECULE TYPE: protein

; HYPOTHETICAL: YES

US-08-788-231A-15

Query Match 90.0%; Score 36; DB 3; Length 462;

Best Local Similarity 100.0%; Pred. No. 3.3;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VNMAEGD 8

Db 291 VNMAEGD 297

RESULT 3

US-08-670-479-18

; Sequence 18, Application US/08670479

; Patent No. 5973133

; GENERAL INFORMATION:

; APPLICANT: Hardy, John A.

; APPLICANT: Goate, Alison M.

; TITLE OF INVENTION: MUTANT S182 GENES

; NUMBER OF SEQUENCES: 24

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SmithKline Beecham Corporation

; STREET: 709 Swedeland Road

; CITY: King of Prussia

; STATE: PA

; COUNTRY: U.S.A.

US-08-670-479-18

Query Match 90.0%; Score 36; DB 2; Length 463;

Best Local Similarity 100.0%; Pred. No. 3.3;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VNMAEGD 8

Db 292 VNMAEGD 298

RESULT 4

US-08-670-964-4

; Sequence 4, Application US/08670964

; Patent No. 6010874

; GENERAL INFORMATION:

; APPLICANT: Hardy, John A.

; TITLE OF INVENTION: EARLY ONSET ALZHEIMER'S DISEASE

; TITLE OF INVENTION: GENE AND GENE PRODUCTS

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SmithKline Beecham Corporation

; STREET: 709 Swedeland Road - UW2220; P.O. Box 15

; CITY: Philadelphia

; STATE: PA

; COUNTRY: USA

; ZIP: 19406

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/670,964

; FILING DATE: 26-JUN-1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/001,142

FILING DATE: 13-JUL-1995
APPLICATION NUMBER: 60/001,501
FILING DATE: 18-JUL-1995
ATTORNEY/AGENT INFORMATION:
NAME: Han, William T
REGISTRATION NUMBER: 34,344
REFERENCE/DOCKET NUMBER: P50358
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5219
TELEFAX: 610-270-5090
TELEX:

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 463 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-670-964-4

Query Match 90.0%; Score 36; DB 3; Length 463;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 VNMAEGD 8
Db 292 VNMAEGD 298

RESULT 5

US-08-888-077A-4
Sequence 4, Application US/08888077A
Patent No. 6020143

GENERAL INFORMATION:

APPLICANT: ST. GEORGE-HYSLOP, PETER H
APPLICANT: ROMMENS, JOHANNA M
APPLICANT: FRASER, PAUL E
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
TO ALZHEIMER'S DISEASE AND USES THEREFOR.
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: LERNER, DAVID, LITTENBERG, KRUMHOLZ & MENTLIK
STREET: 600 SOUTH AVENUE WEST
CITY: WESTFIELD
STATE: NJ
COUNTRY: USA
ZIP: 07090-1497

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/888,077A
FILING DATE: 03-JUL-1997
CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/592,541
FILING DATE: 26-JAN-1996
ATTORNEY/AGENT INFORMATION:

NAME: PALISI, THOMAS M

REGISTRATION NUMBER: 36,629
REFERENCE/DOCKET NUMBER: SCHERING 3.0-017 CIP CIP IV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 654-5000
TELEFAX: (908) 654-7866

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:
LENGTH: 463 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-888-077A-4

Query Match 90.0%; Score 36; DB 3; Length 463;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 VNMAEGD 8
Db 292 VNMAEGD 298

RESULT 6

US-08-788-231A-17
Sequence 17, Application US/08788231A
Patent No. 6019974

GENERAL INFORMATION:

APPLICANT: L'Hernault, Steven W.
TITLE OF INVENTION: SPE-4 RELATED PEPTIDES, ANTIBODIES AND
METHODS
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: US
ZIP: 80303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/788,231A
FILING DATE: 24-JAN-1997
CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/010,672
FILING DATE: 26-JAN-1996

ATTORNEY/AGENT INFORMATION:

NAME: Ferber, Donna M.
REGISTRATION NUMBER: 33,878
REFERENCE/DOCKET NUMBER: 60-95

TELECOMMUNICATION INFORMATION:

TELEPHONE: (303) 499-8080
TELEFAX: (303) 499-8089

INFORMATION FOR SEQ ID NO: 17:

SEQUENCE CHARACTERISTICS:

LENGTH: 465 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: not relevant
MOLECULE TYPE: protein
HYPOTHETICAL: YES

US-08-788-231A-17

Query Match 90.0%; Score 36; DB 3; Length 465;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 VNMAEGD 8
Db 294 VNMAEGD 300

RESULT 7

US-08-967-101-2
Sequence 2, Application US/08967101
Patent No. 5840540

GENERAL INFORMATION:

APPLICANT: ST. GEORGE-HYSLOP, PETER H

APPLICANT: ROMMENS, JOHANNA M
APPLICANT: FRASER, PAUL E
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
TO ALZHEIMER'S DISEASE
NUMBER OF SEQUENCES: 183
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: High Street Tower - 125 High Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/967,101
FILING DATE: 10-NOV-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/592,541
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pitcher, Edmund R.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 467 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-967-101-2

Query Match 90.0%; Score 36; DB 2; Length 467;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VNMAEGD 8
Db 296 VNMAEGD 302

RESULT 8
US-08-967-101-4
Sequence 4, Application US/08967101
Patent No. 5840540
GENERAL INFORMATION:
APPLICANT: ST. GEORGE-HYSLOP, PETER H
APPLICANT: ROMMENS, JOHANNA M
APPLICANT: FRASER, PAUL E
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
TO ALZHEIMER'S DISEASE
NUMBER OF SEQUENCES: 183
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: High Street Tower - 125 High Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: 08/592,541
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pitcher, Edmund R.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 134:
SEQUENCE CHARACTERISTICS:
LENGTH: 467 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-967-101-134

FILING DATE: 10-NOV-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/592,541
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pitcher, Edmund R.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 467 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-967-101-4

Query Match 90.0%; Score 36; DB 2; Length 467;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VNMAEGD 8
Db 296 VNMAEGD 302

RESULT 9
US-08-967-101-134
Sequence 134, Application US/08967101
Patent No. 5840540
GENERAL INFORMATION:
APPLICANT: ST. GEORGE-HYSLOP, PETER H
APPLICANT: ROMMENS, JOHANNA M
APPLICANT: FRASER, PAUL E
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
TO ALZHEIMER'S DISEASE
NUMBER OF SEQUENCES: 183
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: High Street Tower - 125 High Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/967,101
FILING DATE: 10-NOV-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/592,541
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pitcher, Edmund R.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 134:
SEQUENCE CHARACTERISTICS:
LENGTH: 467 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-967-101-134

Query Match 90.0%; Score 36; DB 2; Length 467;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VNMAEGD 8
Db 296 VNMAEGD 302

RESULT 10
US-08-592-541-2
; Sequence 2, Application US/08592541
; Patent No. 5986054
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
; APPLICANT: ROMMENS, JOHANNA M
; APPLICANT: FRASER, PAUL E
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
; TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 183
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: High Street Tower - 125 High Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,541
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Pitcher, Edmund R.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 467 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-592-541-2

Query Match 90.0%; Score 36; DB 2; Length 467;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VNMAEGD 8
Db 296 VNMAEGD 302

RESULT 11
US-08-592-541-4
; Sequence 4, Application US/08592541
; Patent No. 5986054
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
; APPLICANT: ROMMENS, JOHANNA M
; APPLICANT: FRASER, PAUL E
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
; TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 183
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: High Street Tower - 125 High Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,541
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Pitcher, Edmund R.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 467 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-592-541-4

Query Match 90.0%; Score 36; DB 2; Length 467;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VNMAEGD 8
Db 296 VNMAEGD 302

RESULT 12
US-08-592-541-134
; Sequence 134, Application US/08592541
; Patent No. 5986054
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
; APPLICANT: ROMMENS, JOHANNA M
; APPLICANT: FRASER, PAUL E
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
; TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 183
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: High Street Tower - 125 High Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,541
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Pitcher, Edmund R.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 134:
; SEQUENCE CHARACTERISTICS:

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VNMAEGD 8
 Db 296 VNMAEGD 302

RESULT 14

US-08-670-964-2
 ; Sequence 2, Application US/08670964
 ; Patent No. 6010874
 ; GENERAL INFORMATION:
 ; APPLICANT: Hardy, John A.
 ; TITLE OF INVENTION: EARLY ONSET ALZHEIMER'S DISEASE
 ; TITLE OF INVENTION: GENE AND GENE PRODUCTS
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SmithKline Beecham Corporation
 ; STREET: 709 Swedeland Road - UW2220; P.O. Box 15
 ; CITY: Philadelphia
 ; STATE: PA
 ; COUNTRY: USA
 ; ZIP: 19406
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/670,964
 ; FILING DATE: 26-JUN-1996
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 60/001,142
 ; FILING DATE: 13-JUL-1995
 ; APPLICATION NUMBER: 60/001,501
 ; FILING DATE: 18-JUL-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Hag, William T
 ; REGISTRATION NUMBER: 34,344
 ; REFERENCE/DOCKET NUMBER: P50358
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 610-270-5219
 ; TELEFAX: 610-270-5090
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 467 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-670-964-2

Query Match 90.0%; Score 36; DB 3; Length 467;
 Best Local Similarity 100.0%; Pred. No. 3.3;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VNMAEGD 8
 Db 296 VNMAEGD 302

RESULT 15

US-08-888-077A-2
 ; Sequence 2, Application US/08888077A
 ; Patent No. 6020143
 ; GENERAL INFORMATION:
 ; APPLICANT: ST. GEORGE-HYSLOP, PETER H
 ; APPLICANT: ROMMENS, JOHANNA M
 ; APPLICANT: FRASER, PAUL E
 ; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED

Query Match 90.0%; Score 36; DB 3; Length 467;
 Best Local Similarity 100.0%; Pred. No. 3.3;

LENGTH: 467 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-592-341-134

Query Match 90.0%; Score 36; DB 2; Length 467;
 Best Local Similarity 100.0%; Pred. No. 3.3;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VNMAEGD 8
 Db 296 VNMAEGD 302

RESULT 13

US-08-923-454A-10
 ; Sequence 10, Application US/08923454A
 ; Patent No. 6004794
 ; GENERAL INFORMATION:
 ; APPLICANT: Creasy, Caretha
 ; APPLICANT: Livi, George
 ; APPLICANT: Karkan, Eric
 ; APPLICANT: Clinkenbeard, Helen
 ; APPLICANT: Browne, Michael
 ; APPLICANT: Southan, Christopher
 ; TITLE OF INVENTION: HUMAN SERINE PROTEASE
 ; NUMBER OF SEQUENCES: 40
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SmithKline Beecham Corporation
 ; STREET: 709 Swedeland Road
 ; CITY: King of Prussia
 ; STATE: PA
 ; COUNTRY: USA
 ; ZIP: 19406
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq Version 1.5
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/923,454A
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 60/025436
 ; FILING DATE: 06-SEPT-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Baumeister, Kirk
 ; REGISTRATION NUMBER: 33,833
 ; REFERENCE/DOCKET NUMBER: P50547
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 610-270-5096
 ; TELEFAX: 610-270-5090
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 10:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 467 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; FRAGMENT TYPE: N-terminal
 ; ORIGINAL SOURCE:
 ; US-08-923-454A-10

;; TITLE OF INVENTION: TO ALZHEIMER'S DISEASE AND USES THEREFOR.
;; NUMBER OF SEQUENCES: 41
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: LERNER, DAVID, LITTENBERG, KRUMHOLZ & MENTLIK
;; STREET: 600 SOUTH AVENUE WEST
;; CITY: WESTFIELD
;; STATE: NJ
;; COUNTRY: USA
;; ZIP: 07090-1497
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: ASCII
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/888,077A
;; FILING DATE: 03-JUL-1997
;; CLASSIFICATION: 530
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/592,541
;; FILING DATE: 26-JAN-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: PALISI, THOMAS M
;; REGISTRATION NUMBER: 36,629
;; REFERENCE/DOCKET NUMBER: SCHERING 3.0-017 CIP CIP IV
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (908) 654-5000
;; TELEFAX: (908) 654-7866
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 467 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-888-077A-2

Query Match 90.0%; Score 36; DB 3; Length 467;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2 VNMAEGD 8
|||||||
Db 296 VNMAEGD 302

Search completed: September 6, 2001, 16:39:35
Job time: 124 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:45:48 ; Search time 134.15 Seconds
(without alignments)
4.543 Million cell updates/sec

Title: US-09-603-713-6
Perfect score: 40
Sequence: 1 SVNNAEGD 8
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	36	90.0	374	2 S63684	presenilin 1, spli
2	36	90.0	463	2 S63683	presenilin I-463 -
3	36	90.0	463	2 JC5081	presenilin 1 prote
4	36	90.0	467	2 S58396	presenilin 1, spli
5	36	90.0	467	2 JC5080	presenilin 1 prote
6	36	90.0	467	2 I78388	S182 protein - mou
7	34	85.0	562	2 H70688	hypothetical prote
8	33	82.5	433	2 JC5390	presenilin-alpha -
9	32	80.0	354	2 T39599	conserved hypothet
10	32	80.0	1064	2 F86182	hypothetical prote
11	31	77.5	300	2 A84399	hypothetical prote
12	31	77.5	679	1 B48127	dnak-type molecula
13	31	77.5	679	1 I56581	dnak-type molecula
14	31	77.5	679	2 A48127	dnak-type molecula
15	31	77.5	849	2 E82038	GGDEF family prote
16	30	75.0	233	2 JH0372	42K surface glycop
17	30	75.0	334	2 T21562	hypothetical prote
18	30	75.0	386	2 B71407	hypothetical prote
19	30	75.0	510	2 A70580	probable UDP-N-ace
20	29	72.5	152	2 S57425	cysteine proteinase
21	29	72.5	274	2 H83525	conserved hypothet
22	29	72.5	278	2 S06173	development-specif
23	29	72.5	279	2 T25555	hypothetical prote
24	29	72.5	334	2 T51245	COP1-interacting p
25	29	72.5	389	2 G84245	NADH dehydrogenase
26	29	72.5	405	2 S75974	hypothetical prote
27	29	72.5	406	2 I40373	ornithine acetyltr
28	29	72.5	481	2 H69593	6-phospho-beta-glu
29	29	72.5	765	2 G64502	hypothetical prote

30 29 72.5 810 2 A64742 hypothetical prote
31 29 72.5 810 2 C85502 hypothetical prote
32 29 72.5 829 2 E64114 translation initia
33 29 72.5 1420 2 T02644 ABC-type transport
34 28 70.0 138 1 D69979 conserved hypothet
35 28 70.0 184 2 T25584 hypothetical prote
36 28 70.0 187 2 S69315 hypothetical prote
37 28 70.0 286 1 RLPUGG rRNA N-glycosidase
38 28 70.0 313 2 S60713 polygalacturonase-
39 28 70.0 330 2 H64894 tellurite resistan
40 28 70.0 330 2 E85735 tellurite resistan
41 28 70.0 335 2 G72304 histidinol-phospha
42 28 70.0 342 2 A45966 Ig alpha chain C r
43 28 70.0 342 2 S23764 polygalacturanase-
44 28 70.0 344 1 AHMS Ig alpha chain C r
45 28 70.0 346 2 S57183 hypothetical prote

ALIGNMENTS

RESULT 1
S63684
presenilin 1, splice form 374 - human
N:Alternate names: Alzheimer's disease protein 3
C:Species: Homo sapiens (man)
C>Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 29-Sep-1999
C:Accession: S63684
R:Sahara, N.; Yahagi, Y.; Takagi, H.; Kondo, T.; Okochi, M.; Usami, M.; Shirasawa, T.
FEBS Lett. 381, 7-11, 1996
A:Title: Identification and characterization of presenilin I-467, I-463 and I-374.
A:Reference number: S63683; MUID:96193901
A:Accession: S63684
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-374 <SAH>
A:Cross-references: EMBL:U40380; NID:gl244639; PIDN:AAB05895.1; PID:gl244640
C:Genetics:
A:Gene: GDB:PSEN1; AD3; FAD; S182; PSI
A:Cross-references: GDB:135682; OMIM:104311
A:Map position: 14q24.3-14q24.3
C:Superfamily: presenilin

Query Match 90.0%; Score 36; DB 2; Length 374;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VNNAEGD 8
DB 292 VNNAEGD 298

RESULT 2
S63683
presenilin I-463 - human
C:Species: Homo sapiens (man)
C>Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 29-Sep-1999
C:Accession: S63683
R:Sahara, N.; Yahagi, Y.; Takagi, H.; Kondo, T.; Okochi, M.; Usami, M.; Shirasawa, T.
FEBS Lett. 381, 7-11, 1996
A:Title: Identification and characterization of presenilin I-467, I-463 and I-374.
A:Reference number: S63683; MUID:96193901
A:Accession: S63683
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-463 <SAH>
A:Cross-references: EMBL:U40379; NID:gl244637; PIDN:AAB05894.1; PID:gl244638
C:Superfamily: presenilin

Query Match 90.0%; Score 36; DB 2; Length 463;
Best Local Similarity 100.0%; Pred. No. 3.3;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VNMAEGD 8
 Db 292 VNMAEGD 298

RESULT 3

JC5081
 presenilin 1 protein isoform 463 - lesser mouse lemur
 C:Species: Microcebus murinus (lesser mouse lemur)
 C:Date: 31-Jan-1997 #sequence_revision 31-Jan-1997 #text_change 13-Sep-1998
 C:Accession: JC5081
 R:Calenda, A.; Mestres-Frances, N.; Czech, C.; Pradier, L.; Petter, A.; Bons, N.; Bellis, Biochem. Biophys. Res. Commun. 228, 430-439, 1996
 A:Title: Molecular cloning, sequencing, and brain expression of the presenilin 1 gene in A:Reference number: JC5080; MUID:97079199

A:Contents: brain
 A:Accession: JC5081
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-463 <CAL>
 A:Cross-references: EMBL:Z71333
 C:Comment: This protein is an intermembrane protein with seven transmembrane domains. It C:Genetics:
 A:Gene: psi
 A:Map position: 14
 C:Superfamily: presenilin
 C:Keywords: transmembrane protein
 F:78-96/Domain: transmembrane #status predicted <TM1>
 F:129-150/Domain: transmembrane #status predicted <TM2>
 F:160-181/Domain: transmembrane #status predicted <TM3>
 F:191-209/Domain: transmembrane #status predicted <TM4>
 F:217-234/Domain: transmembrane #status predicted <TM5>
 F:240-257/Domain: transmembrane #status predicted <TM6>
 F:404-424/Domain: transmembrane #status predicted <TM7>

Query Match 90.0%; Score 36; DB 2; Length 463;
 Best Local Similarity 100.0%; Pred. No. 3.3;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VNMAEGD 8
 Db 292 VNMAEGD 298

RESULT 4

S58396
 presenilin 1, splice form 467 - human
 N:Alternate names: Alzheimer's disease protein 3; protein S182
 C:Species: Homo sapiens (man)
 C:Date: 29-Jan-1998 #sequence_revision 13-Feb-1998 #text_change 29-Sep-1999
 C:Accession: S58396; S71401; S71402
 R:Sherrington, R.; Rogeev, E.I.; Liang, Y.; Roaeva, E.A.; Levesque, G.; Ikeda, M.; Chiro, I.; Pinessi, L.; Nee, L.; Chumakov, I.; Pollen, D.; Brookes, A.; Sanseau, P.; Pollen, E.; Rommens, J.M.; St George-Hyslop, P.H.
 Nature 375, 754-760, 1995
 A:Title: Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's A:Reference number: S58095; MUID:95319502
 A:Accession: S58396
 A:Molecule type: mRNA
 A:Residues: 1-467 <SHE>
 A:Cross-references: EMBL:L42110; NID:g904118; PIDN:AA846416.1; PID:g904119
 R:Vidal, R.; Ghiso, J.; Wisniewski, T.; Frangione, B.
 FEBS Lett. 393, 19-23, 1996
 A:Title: Alzheimer's presenilin 1 gene expression in platelets and megakaryocytes. Ident A:Reference number: S71401; MUID:96397521
 A:Accession: S71401
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 24-32;254-256,290-292;316-317,376-379 <VIW>
 A:Experimental source: Dami megakaryotic cell line (ATCC CRL-9792) and platelets

C:Genetics:
 A:Gene: GDB:PSEN1; AD3; PAD; S182; PS1
 A:Cross-references: GDB:135682; OMIM:104311
 A:Map position: 14q24.3-14q24.3
 C:Superfamily: presenilin
 C:Keywords: alternative splicing; Alzheimer's disease; glycoprotein; transmembrane pr F:82-100/Domain: transmembrane #status predicted <TM1>
 F:133-154/Domain: transmembrane #status predicted <TM2>
 F:164-185/Domain: transmembrane #status predicted <TM3>
 F:195-213/Domain: transmembrane #status predicted <TM4>
 F:221-238/Domain: transmembrane #status predicted <TM5>
 F:244-264/Domain: transmembrane #status predicted <TM6>
 F:281-301/Domain: transmembrane #status predicted <TM7>
 F:408-428/Domain: transmembrane #status predicted <TM8>
 F:433-453/Domain: transmembrane #status predicted <TM9>
 F:279-405/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 90.0%; Score 36; DB 2; Length 467;
 Best Local Similarity 100.0%; Pred. No. 3.4;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VNMAEGD 8
 Db 296 VNMAEGD 302

RESULT 5

JC5080
 presenilin 1 protein isoform 467 - lesser mouse lemur
 C:Species: Microcebus murinus (lesser mouse lemur)
 C:Date: 31-Jan-1997 #sequence_revision 31-Jan-1997 #text_change 20-Jun-2000
 C:Accession: JC5080
 R:Calenda, A.; Mestres-Frances, N.; Czech, C.; Pradier, L.; Petter, A.; Bons, N.; Bell Biochem. Biophys. Res. Commun. 228, 430-439, 1996
 A:Title: Molecular cloning, sequencing, and brain expression of the presenilin 1 gene A:Reference number: JC5080; MUID:97079199
 A:Accession: JC5080
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-467 <CAL>
 A:Cross-references: EMBL:Z71333; NID:g1707591; PIDN:CAA95930.1; PID:g1707592
 A:Experimental source: brain
 C:Comment: This protein is an integral membrane protein with seven transmembrane dom C:Genetics:
 A:Gene: psi
 A:Map position: 14
 C:Superfamily: presenilin
 C:Keywords: transmembrane protein
 F:82-100/Domain: transmembrane #status predicted <TM1>
 F:133-154/Domain: transmembrane #status predicted <TM2>
 F:164-185/Domain: transmembrane #status predicted <TM3>
 F:195-213/Domain: transmembrane #status predicted <TM4>
 F:221-238/Domain: transmembrane #status predicted <TM5>
 F:244-261/Domain: transmembrane #status predicted <TM6>
 F:408-428/Domain: transmembrane #status predicted <TM7>

Query Match 90.0%; Score 36; DB 2; Length 467;
 Best Local Similarity 100.0%; Pred. No. 3.4;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VNMAEGD 8
 Db 296 VNMAEGD 302

RESULT 6

S182 protein - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 27-Feb-1997 #sequence_revision 27-Feb-1997 #text_change 29-Sep-1999
 C:Accession: I78388

R:Sherrington, R.; Rogaev, E.I.; Liang, Y.; Rogaeva, E.A.; Levesque, G.; Ikeda, M.; Chi,
 ero, I.; Pinessi, L.; Nee, L.; Chumakov, I.; Pollen, D.; Brookes, A.; Sanseau, P.; Pollen,
 E.; Rommens, J.M.; St George-Hyslop, P.H.

Nature 375, 754-760, 1995
 A:Title: Cloning of a gene bearing missense mutations in early-onset familial Alzheimer
 A:Reference number: 158095; MUID:95319502

A:Accession: 178388

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1-467 <RES>

A:Cross-references: GB:L42177; NID:g904129; PIDN:AAC42094.1; PID:g904130

C:Superfamily: presenilin

Query Match 90.0%; Score 36; DB 2; Length 467;

Best Local Similarity 100.0%; Pred. No. 3.4;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VNMAEGD 8

|||||

Db 296 VNMAEGD 302

RESULT 7

H70688

Hypothetical protein Rv2797c - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: H70688

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

Rejandream, M.A.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Nature 393, 537-544, 1998

A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987

A:Accession: H70688

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-562 <COL>

A:Cross-references: GB:Z81331; GB:AL123456; NID:g3261650; PIDN:CAB03649.1; PID:e276819;

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: Rv2797c

Query Match 85.0%; Score 34; DB 2; Length 562;

Best Local Similarity 75.0%; Pred. No. 12;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVNMAEGD 8

|||||

Db 158 AVNMADGD 165

RESULT 8

JC5390

presenilin-alpha - African clawed frog

C:Species: Xenopus laevis (African clawed frog)

C:Date: 04-Jun-1997 #sequence_revision 18-Jul-1997 #text_change 20-Jun-2000

C:Accession: JC5390

R:Tsujiyama, A.; Yasojima, K.; Hashimoto-Gotoh, T.

Biochem. Biophys. Res. Commun. 231, 392-396, 1997

A:Title: Cloning of Xenopus presenilin-alpha and -beta cDNAs and their differential exp

A:Reference number: JC5390; MUID:97223465

A:Accession: JC5390

A:Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-433 <TSU>

A:Cross-references: DDBJ:D84427; NID:g1944353; PIDN:BAA19570.1; PID:g1944354

A:Experimental source: brain

C:Comment: This protein plays a role in negative regulation of apoptotic cascades during

C:Superfamily: presenilin

F:48-66/Domain: transmembrane #status predicted <TM1>
 F:99-119/Domain: transmembrane #status predicted <TM2>
 F:130-179/Domain: transmembrane #status predicted <TM3>
 F:161-178/Domain: transmembrane #status predicted <TM4>
 F:187-203/Domain: transmembrane #status predicted <TM5>
 F:210-227/Domain: transmembrane #status predicted <TM6>
 F:374-394/Domain: transmembrane #status predicted <TM7>

Query Match 82.5%; Score 33; DB 2; Length 433;

Best Local Similarity 85.7%; Pred. No. 15;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 VNMAEGD 8

|||||

Db 262 VNMAEGD 268

RESULT 9

T39599

conserved hypothetical protein SPBC16G5.07c - fission yeast (Schizosaccharomyces pombe

C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 18-Feb-2000

C:Accession: T39599

R:Lyne, M.; Wood, V.; Rajandream, M.A.; Barrell, B.G.; Beck, A.; Reinhardt, R.

submitted to the EMBL Data Library, May 1998

A:Reference number: Z21866

A:Accession: T39599

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-334 <LYN>

A:Cross-references: EMBL:AL023554; PIDN:CAA19027.1; GSPDB:GN00067; SPDB:SPBC16G5.07c

A:Experimental source: strain 972h; cosmid c16G5

C:Genetics:

A:Gene: SPDB:SPBC16G5.07c

A:Map position: 2 265/2

A:Introns: 72/2; 265/2

C:Superfamily: erythrocyte band 7 integral membrane protein

Query Match 80.0%; Score 32; DB 2; Length 354;

Best Local Similarity 62.5%; Pred. No. 20;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVNMAEGD 8

|||||

Db 229 AINVAEGD 236

RESULT 10

F86182

hypothetical protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: F86182

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon

Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar,

ansen, N.P.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: F86182

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1064 <STO>

A:Cross-references: GB:AE005172; NID:g7211974; PIDN:AAF40445.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

Fri Sep 7 10:58:43 2001

C;Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology

Query Match 80.0%; Score 32; DB 2; Length 1064;
 Best Local Similarity 75.0%; Pred. No. 66;
 Matches 6; Conservative 1; Mismatches 0; Gaps 0;

QY 1 SVNMAEGD 8

DB 32 SVNLRGEGD 39

RESULT 11

A84399 hypothetical protein Vng2488c [imported] - Halobacterium sp. NRC-1

C;Species: Halobacterium sp. NRC-1
 C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C;Accession: A84399

R;Ng W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabc

Jung, K.H.; Alam, M.; Freitas, T.

Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000

A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omar, A.D.; Ebhardt, H.; Lowe, T.M.; Li

A;Title: Genome sequence of Halobacterium species NRC-1.

A;Reference number: A84160; MUID:20504483

A;Accession: A84399

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-300 <STO>

A;Cross-references: GB:AE004437; NID:g10581883; PIDN:AAG20557.1; GSPDB:GN00138

C;Genetics:

A;Gene: VNG2488C

Query Match 77.5%; Score 31; DB 2; Length 300;
 Best Local Similarity 75.0%; Pred. No. 28;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVNMAEGD 8

DB 42 SVDVAEGD 49

RESULT 12

B48127 dnak-type molecular chaperone precursor, mitochondrial - human

N;Alternate names: heat shock protein 70 homolog; PBP74; peptide-binding protein 74

C;Species: Homo sapiens (man)

C;Date: 21-Jan-1994 #sequence_revision 26-May-1995 #text_change 20-Apr-2000

C;Accession: B48127; A55623

R;Domancic, S.Z.; Denagel, D.C.; Dahlseid, J.N.; Green, J.M.; Pierce, S.K.

Mol. Cell. Biol. 13, 3598-3610, 1993

A;Title: Cloning of the gene encoding peptide-binding protein 74 shows that it is a new

A;Reference number: A48127; MUID:93268309

A;Accession: B48127

A;Molecule type: mRNA

A;Residues: 1-679 <DOM>

A;Cross-references: GB:L11066; NID:g307322

A;Note: sequence extracted from NCBI backbone (NCBIN:132585, NCBI:132586)

A;Note: this ORF is not annotated in Genbank entry HMPBP, release 106

R;Bhattacharya, T.; Karnezis, A.N.; Murphy, S.P.; Freeman, B.C.; Phillips, B.

J. Biol. Chem. 270, 1705-1710, 1995

A;Title: Cloning and subcellular localization of human mitochondrial hsp70.

A;Reference number: A55623; MUID:95130547

A;Accession: A55623

A;Status: not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1-539, 'R', 541-679 <BHA>

A;Cross-references: GB:L15189; NID:g292058; PIDN:AAA67526.1; PID:g292059

A;Note: authors definitively demonstrate mitochondrial localization

C;Genetics:

A;Gene: GDB:HSPA9B; HSPA9; GRP75; PBP74; MOT2; MTHSP75

A;Cross-references: GDB:626151; OMIM:600548

A;Map position: 5q31.1-5q31.1

C;Function:

A;Description: involved in protein folding and assembling/disassembling of protein co

C;Superfamily: heat shock protein 70

C;Keywords: ATP; heat shock; mitochondrion; molecular chaperone; stress-induced prote

F;1-46/Domain: transit peptide (mitochondrion) #status predicted <TP>

F;47-679/Product: mitochondrial hsp70 #status predicted <MAT>

Query Match 77.5%; Score 31; DB 1; Length 679;
 Best Local Similarity 85.7%; Pred. No. 68;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVNMAEG 7

DB 581 AVNMAEG 587

RESULT 13

I56581 dnak-type molecular chaperone grp75 precursor - rat

N;Alternate names: glucose regulated protein 75; grp75; heat shock protein 70 homolog

C;Species: Rattus norvegicus (Norway rat)

C;Date: 26-Jul-1996 #sequence_revision 11-Apr-1997 #text_change 16-Jul-1999

C;Accession: I56581; B35488; I53019

R;Massa, S.M.; Longo, F.M.; Zuo, J.; Wang, S.; Chen, J.; Sharp, F.R.

J. Neurosci. Res. 40, 807-819, 1995

A;Title: Cloning of rat grp75, an hsp70-family member, and its expression in normal a

A;Reference number: I56581; MUID:95356254

A;Accession: I56581

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-679 <RES>

A;Cross-references: GB:S78556; NID:g1000438; PIDN:AAB34982.1; PID:g1000439

A;Experimental source: strain Sprague-Dawley, brain

R;Akamizu, T.; Saiji, M.; Kohn, L.D.

Biochem. Biophys. Res. Commun. 170, 351-358, 1990

A;Title: A microsequencing approach to identify proteins which appear to interact wit

A;Reference number: A35488; MUID:90321251

A;Accession: B35488

A;Molecule type: protein

A;Residues: 80-98; 484-486, 'T', 488-492, 'Q', 494-503 <AKA>

R;Webster, T.J.; Naylor, D.J.; Hartman, D.J.; Hoj, P.B.; Hoogenraad, N.J.

DNA Cell Biol. 13, 1213-1220, 1994

A;Title: cDNA cloning and efficient mitochondrial import of pre-mTHSP70 from rat live

A;Reference number: I53019; MUID:95110439

A;Accession: I53019

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-36, 'V', 38-80, 'S', 82-372, 'R', 374-588, 'V', 590-679 <RE2>

A;Cross-references: GB:S75280; NID:g896231; PIDN:AAB33049.1; PID:g896232

C;Function:

A;Description: involved in protein folding and assembling/disassembling of protein co

C;Superfamily: heat shock protein 70

C;Keywords: ATP; mitochondrion; molecular chaperone; stress-induced protein

F;1-46/Domain: transit peptide (mitochondrion) #status predicted <TP>

F;47-679/Product: mitochondrial heat shock protein 70 homolog #status predicted <MAT>

Query Match 77.5%; Score 31; DB 1; Length 679;
 Best Local Similarity 85.7%; Pred. No. 68;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVNMAEG 7

DB 581 AVNMAEG 587

RESULT 14

A48127

dnak-type molecular chaperone precursor, mitochondrial - mouse

N;Alternate names: C3H strain-specific antigenic protein (CSA); heat shock protein 70

C;Species: Mus musculus (house mouse)

C:Date: 21-Jan-1994 #sequence_revision 23-Mar-1995 #text_change 20-Jun-2000
C:Accession: S39839; A48127; JN0808; A45497; B48698; A48698
R:Michikawa, Y.; Baba, T.; Arai, Y.; Sakakura, T.; Kusakabe, M.
FEBS Lett. 336, 27-33, 1993
A:Title: Structure and organization of the gene encoding a mouse mitochondrial stress-70
A:Reference number: S39839; MUID:94085585
A:Accession: S39839
A:Molecule type: DNA
A:Residues: 1-679 <MIC>
A:Cross-references: GB:D17666; NID:9460726; PIDN:BAA04548.1; PID:g903309
R:Domianico, S.Z.; DeNagel, D.C.; Dahlseid, J.N.; Green, J.M.; Pierce, S.K.
Mol. Cell. Biol. 13, 3598-3610, 1993
A:Title: Cloning of the gene encoding peptide-binding protein 74 shows that it is a new
A:Reference number: A48127; MUID:93268309
A:Accession: A48127
A:Molecule type: mRNA
A:Residues: 1-521 'S', 523-679 <DOM>
A:Note: sequence extracted from NCBI backbone (NCBIN:132583, NCBIP:132584)
A:Accession: C48127
A:Molecule type: protein
A:Residues: 47-63; 87-99; 188-202; 219-237; 266-284; 542-563 <DO2>
A:Note: antibody to amino end of this protein suggested localization to cytoplasmic vesicle
R:Michikawa, Y.; Baba, T.; Arai, Y.; Sakakura, T.; Tanaka, M.; Kusakabe, M.
Biochem. Biophys. Res. Commun. 196, 223-232, 1993
A:Title: Antigenic protein specific for C3H strain mouse is a mitochondrial stress-70 protein
A:Reference number: JN0808; MUID:94029998
A:Accession: JN0808
A:Molecule type: mRNA
A:Residues: 1-617 'V', 619-623 'R', 625-679 <MI2>
A:Cross-references: DDBJ:D17556; NID:9407340; PIDN:BAA04493.1; PID:g407341
R:Wadhwa, R.; Kaul, S.C.; Ikawa, Y.; Sugimoto, Y.
J. Biol. Chem. 268, 6615-6621, 1993
A:Title: Identification of a novel member of mouse hsp70 family. Its association with cell
A:Reference number: A45497; MUID:93203261
A:Accession: A45497
A:Molecule type: mRNA
A:Residues: 1-122 'I', 124-617 'V', 619-623 'R', 625-679 <WAD>
A:Cross-references: GB:D11089; NID:g91763; PIDN:BAA01862.1; PID:g9391764
A:Note: sequence extracted from NCBI backbone (NCBIN:127943, NCBIP:127944)
R:Wadhwa, R.; Kaul, S.C.; Sugimoto, Y.; Mitsui, Y.
J. Biol. Chem. 268, 22239-22242, 1993
A:Title: Induction of cellular senescence by transfection of cytosolic mortalin cDNA in
A:Reference number: A48698; MUID:94042962
A:Accession: A48698
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-122 'I', 124-202 'D', 204-679 <WAZ>
A:Experimental source: NIH 3T3 cells
A:Note: this form designated perinuclear form (mot-2) from an immortal cell line
A:Note: sequence extracted from NCBI backbone (NCBIP:138937)
A:Accession: A48698
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-122 'I', 124-202 'D', 204-617 'V', 619-623 'R', 625-679 <WAZ>
A:Experimental source: C61-ICR embryonic fibroblasts
A:Note: this form designated cytoplasmic form (mot-1) from a mortal cell line
A:Note: sequence extracted from NCBI backbone (NCBIP:138936)
C:Comment: The human homolog of this protein has been shown definitively to be a mitochondrial
C:Genetics:
A:Genome: nuclear
C:Function:
A:Description: Involved in protein folding and assembling/disassembling of protein complex
C:Superfamily: heat shock protein 70
C:Keywords: ATP; mitochondrion; molecular chaperone
F1-45/Domain: transit peptide (mitochondrion) #status predicted <TPP>
F:47-679/Product: dnaK-type molecular chaperone, mitochondrial #status predicted <MAT>

Query Match 77.5%; Score 31; DB 2; Length 679;
Best Local Similarity 85.7%; Pred. No. 68;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVNMAEG 7
DB 581 AVNMAEG 587
RESULT 15
E82038
GGDEF family protein VC2750 [Imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: E82038
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.
charlson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers
I., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833
A:Accession: E82038
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-849 <HEI>
A:Cross-references: GB:AE004340; GB:AE003852; NID:g9657344; PIDN:AAF95889.1; GSPDB:GN
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC2750
A:Map position: 1

Query Match 77.5%; Score 31; DB 2; Length 849;
Best Local Similarity 75.0%; Pred. No. 87;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVNMAEGD 8
DB 330 SVSIARGD 337

Search completed: September 6, 2001, 16:45:49
Job time: 493 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:51:05 ; Search time 72.75 Seconds
(without alignments)
3.767 Million cell updates/sec

Title: US-09-603-713-6
Perfect score: 40
Sequence: 1 SVNNAEGD 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues
Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	90.0	467	1 PSN1_HUMAN	P49768 homo sapien
2	36	90.0	467	1 PSN1_MOUSE	P79802 microcebus
3	36	90.0	467	1 PSN1_MOUSE	P49769 mus musculus
4	36	90.0	468	1 PSN1_RAT	P97887 rattus norv
5	36	90.0	478	1 PSN1_BOVIN	Q9xt97 bos taurus
6	33	82.5	433	1 PSN1_XENLA	O12976 xenopus lae
7	31	77.5	679	1 GR75_CRIGR	O35501 cricetus
8	31	77.5	679	1 GR75_HUMAN	P38646 homo sapien
9	31	77.5	679	1 GR75_HUMAN	P38647 mus musculus
10	31	77.5	679	1 GR75_RAT	P48721 rattus norv
11	30	75.0	233	1 GP42_RAT	P23505 rattus norv
12	30	75.0	510	1 MURF_MYCTU	O06220 mycobacteri
13	30	75.0	674	1 GSPD_VIBCH	P45779 vibrio chol
14	30	75.0	842	1 GYRA_MYCSM	P48354 mycobacteri
15	29	72.5	268	1 YBGF_PSEPU	P43037 pseudomonas
16	29	72.5	278	1 D7_XENLA	P13007 xenopus lae
17	29	72.5	319	1 PP12_ACECL	P48481 acetabulari
18	29	72.5	406	1 ARGJ_BACSU	P36843 b arginine
19	29	72.5	469	1 BGL1_BACSU	P40740 bacillus su
20	29	72.5	810	1 UP05_ECOLI	P39170 escherichia
21	29	72.5	829	1 IF2_HAEIN	P44323 haemophilus
22	28	70.0	138	1 YRRK_BACSU	O34634 bacillus s
23	28	70.0	286	1 RIP1_MOMCH	P16094 momordica c
24	28	70.0	330	1 TEHA_ECOLI	P25396 escherichia
25	28	70.0	342	1 TGIP_PHAVU	P35334 phaseolus v
26	28	70.0	344	1 ALC_MOUSE	P01878 mus musculus
27	28	70.0	346	1 Y9S_YEAST	P47181 saccharomyc
28	28	70.0	415	1 TRSA_STRAM	Q07197 streptomyce
29	28	70.0	419	1 VSV1_TRYCO	P20948 trypanosoma
30	28	70.0	471	1 VSWA_TRYBR	P20946 trypanosoma
31	28	70.0	497	1 DHAL_ASPNG	P41751 aspergillus
32	28	70.0	663	1 TAZ1_SCHPO	P79005 schizosacch
33	28	70.0	822	1 GYRA_STRPN	P72524 streptococc

RESULT 1

ID	PSN1_HUMAN	STANDARD;	PRT;	467 AA.
AC	P49768; Q14762; Q15719; Q15720;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	PRESENILIN 1 (PS-1) (S182 PROTEIN).			
GN	PSN1 OR PSNLI OR AD3 OR PS1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND VARIANTS AD (ISOFORMS I-467 AND I-463).			
RC	TISSUE=Brain;			
RX	MEDLINE=95319502; PubMed=7596406;			
RA	Sherrington R., Rogaeve E.I., Liang Y., Rogaeve E.A., Levesque G., Ikeda M., Chi H., Lin C., Li G., Holman K., Tsuda T., Mar L., Foncin J.-F., Bruni A.C., Montesi M.P., Sorbi S., Rainero I., Pinessi L., Nee L., Chumakov I., Pollen D., Brookes A., Sanseau P., Polinsky R.J., Wasco W., da Silva H.A.R., Haines J.L., Pericak-Vance M.A., Tanzi R.E., Roses A.D., Fraser P.E., Rommens J.M., St George-Hyslop P.H.;			
RA	"Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease.";			
RT	Nature 375:754-760(1995).			
RL	[2]			
RP	SEQUENCE FROM N.A. (ISOFORMS I-463 AND I-374).			
RC	TISSUE=Blood, and Brain;			
RX	MEDLINE=96193901; PubMed=8641442;			
RA	Sahara N., Yahagi Y.-I., Takagi H., Kondo T., Okochi M., Usami M., Shirasawa T., Mori H.;			
RA	"Identification and characterization of presenilin I-467, I-463 and I-374.";			
RL	FEBS Lett. 381:7-11(1996).			
RP	SEQUENCE FROM N.A.			
RA	Rowen L., Madan A., Qin S., Abbasi N., Dors M., Ratcliffe A., Madan A., Dickhoff R., Shaffer T., James R., Lasky S., Hood L.;			
RA	"Complete sequence of the gene for presenilin 1.";			
RL	Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.			
RN	[4]			
RP	SEQUENCE OF 1-113 FROM N.A.			
RA	Tsujiura A., Hashimoto-Gotoh T.;			
RA	Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.			
RL	[5]			
RP	SUBCELLULAR LOCATION, AND TISSUE SPECIFICITY.			
RX	MEDLINE=96160372; PubMed=8574969;			
RA	Kovacs D.M., Faustet H.J., Page K.J., Kim T.-W., Moir R.D., Merriam D.E., Hollister R.D., Hallmark O.G., Mancini R., Felsenstein K.M., Hyman B.T., Tanzi R.E., Wasco W.;			
RA	"Alzheimer-associated presenilins 1 and 2: neuronal expression in brain and localization to intracellular membranes in mammalian cells.";			
RL	Nat. Med. 2:224-229(1996).			

RA REVIEW ON VARIANTS.
 RX MEDLINE=97029239; PubMed=8875251;
 RA Cruts M., Hendriks L., Van Broeckhoven C.;
 RT "The presenilin genes: a new gene family involved in Alzheimer disease
 pathology."; Hum. Mol. Genet. 5:1449-1455(1996).
 RL [17]
 RN
 RP REVIEW ON VARIANTS.
 RX MEDLINE=98180715; PubMed=9521418;
 RA Cruts M., van Broeckhoven C.;
 RT "Presenilin mutations in Alzheimer's disease."; Hum. Mutat. 11:183-190(1998).
 RL [18]
 RN
 RP VARIANTS AD THR-143 AND ALA-384.
 RX MEDLINE=96177673; PubMed=8634711;
 RA Cruts M., Backhovens H., Wang S.-Y., van Gassen G., Theuns J.,
 RA de Jonghe C., Wehnert A., de Voecht J., de Winter G., Cras P.,
 RA Bruylant M., Watson N., Weissenbach J., den Dunnen J.T., Martin J.-J.,
 RA Hendriks L., Van Broeckhoven C.;
 RT "Molecular genetic analysis of familial early-onset Alzheimer's
 disease linked to chromosome 14q24.3."; Hum. Mol. Genet. 4:2363-2372(1995).
 RL [19]
 RN
 RP VARIANTS AD L-82; H-115; T-139; R-163; T-231; L-264; V-392 AND Y-410.
 RX MEDLINE=96177674; PubMed=8634712;
 RA Campion D., Flaman J.-M., Brice A., Hannequin D., Dubois B.,
 RA Martin C., Moreau V., Charbonnier F., Didierjean O., Tardieu S.,
 RA Penet C., Puel M., Pasquier F., le Doze F., Bellis G., Calenda A.,
 RA Heilig R., Martinez M., Mallet J., Bellis M., Clerget-Darpoux F.,
 RA Agid Y., Frebourg T.;
 RT "Mutations of the presenilin I gene in families with early-onset
 Alzheimer's disease."; Hum. Mol. Genet. 4:2373-2377(1995).
 RL [10]
 RN
 RP VARIANTS AD VAL-260; VAL-285 AND VAL-392.
 RX MEDLINE=95379971; PubMed=7551536;
 RA Rogeev E.I., Sherrington R., Rogaeva E.A., Levesque G., Ikeda M.,
 RA Liang Y., Chi H., Lin C., Holman K., Tsuda T., Mar L., Sorbi S.,
 RA Nacias B., Placentini S., Amaducci L., Chumakov I., Cohen D.,
 RA Lannfelt L., Fraser P.E., Romenes J.M., St George-Hyslop P.H.;
 RT "Familial Alzheimer's disease in kindreds with missense mutations in
 a gene on chromosome 1 related to the Alzheimer's disease type 3
 gene."; Nature 376:775-778(1995).
 RL [11]
 RN
 RP VARIANTS AD V-139; V-146; Y-163; T-267; A-280 AND G-280.
 RX MEDLINE=96024664; PubMed=7550356;
 RA Clark R.F., Hutton M., Fuldner R.A., Froelich S., Karan E.,
 RA Talbot C., Crook R., London C., Prihar G., He C., Korenblatt K.,
 RA Martinez A., Wrangé M., Busfield F., Behrens M.I., Myers A., Norton J.,
 RA Morris J., Mehta N., Pearson C., Lincoln S., Baker M., Duff K.,
 RA Zehr C., Perez-Tur J., Houlden H., Ruiz A., Ossa J., Lopera F.,
 RA Arcos M., Madrigal L., Collinge J., Humphreys C., Asworth T.,
 RA Sarter S., Fox N., Harvey R., Kennedy A., Roques P., Cline R.T.,
 RA Phillips C.A., Venter J.C., Forsell L., Axelman K., Lillius L.,
 RA Johnston J., Comburn R., Vitanen M., Winblad B., Kosik K., Haltia M.,
 RA Poyhonen M., Dickson D., Mann D., Neary D., Snowden J., Lantos P.,
 RA Lannfelt L., Rossor M., Roberts G.W., Adams M.D., Hardy J., Goate A.;
 RT "The structure of the presenilin 1 (S182) gene and identification of
 six novel mutations in early onset AD families."; Nat. Genet. 11:219-222(1995).
 RL [12]
 RN
 RP VARIANTS AD PHE-96; ARG-163 AND THR-213.
 RX MEDLINE=96310408; PubMed=8733303;
 RA Kamino K., Sato S., Sakaki Y., Yoshiwa A., Nishiwaki Y., Takeda H.,
 RA Tanabe H., Nishimura T., Li K., St George-Hyslop P.H., Miki T.,
 RA Ogihara T.;
 RT "Three different mutations of presenilin 1 gene in early-onset
 Alzheimer's disease families."; Neurosci. Lett. 208:195-198(1996).
 RL [13]
 RN
 RP VARIANT AD ASP-135.
 RX MEDLINE=97369208; PubMed=92255696;
 RA Crook R., Ellis R., Shanks M., Thal L.J., Perez-Tur J., Baker M.,
 RA Hutton M., Haltia T., Hardy J., Galasko D.;
 RT "Early-onset Alzheimer's disease with a presenilin-1 mutation at the
 RT site corresponding to the Volga German presenilin-2 mutation."; Ann. Neurol. 42:124-128(1997).
 RL [14]
 RN
 RP VARIANT AD ALA-280.
 RX MEDLINE=97442268; PubMed=9298817;
 RA London C.L., Martinez A., Behrens I.M., Kosik K.S., Madrigal L.,
 RA Norton J., Neuman R., Myers A., Busfield F., Wrangé M., Arcos M.,
 RA Arango Viana J.C., Ossa J., Ruiz A., Goate A.M., Lopera F.;
 RT "E280A PS-1 mutation causes Alzheimer's disease but age of onset is
 RT not modified by ApoE alleles."; Hum. Mutat. 10:186-195(1997).
 RL [15]
 RN
 RP VARIANTS AD THR-233 AND THR-278.
 RX MEDLINE=97316242; PubMed=9172170;
 RA Hardy J., Fulham M.J., Nicholson G.A., Stell R.,
 RA St George-Hyslop P.H., Fraser P.E., Kakulas B., Clarnette R.,
 RA Relkin N., Gandy S.E., Schofield P.R., Martins R.N.;
 RT "Two novel (W237F and R278T) presenilin-1 mutations in early-onset
 Alzheimer's disease pedigrees and preliminary evidence for
 RT association of presenilin-1 mutations with a novel phenotype."; NeuroReport 8:1537-1542(1997).
 RL [16]
 RN
 RP VARIANT AD PRO-171.
 RX MEDLINE=99050262; PubMed=9833068;
 RA Ramirez-Duenas M.G., Rogaeva E.A., Leal C.A., Lin C.,
 RA Ramirez-Casillas G.A., Hernandez-Romo J.A., St George-Hyslop P.H.,
 RA Cantu J.M.;
 RT "A novel Leu171Pro mutation in presenilin-1 gene in a Mexican family
 RT with early onset Alzheimer disease."; Ann. Genet. 41:149-153(1998).
 RL [17]
 RN
 RP VARIANT GLY-318.
 RX MEDLINE=99066768; PubMed=9851443;
 RA Mattila K.M., Forsell C., Pirttila T., Rinne J.O., Lehtimäki T.,
 RA Roytta M., Lillius L., Eerola A., St George-Hyslop P.H., Frey H.,
 RA Lannfelt L.;
 RT "The Glu318Gly mutation of the presenilin-1 gene does not necessarily
 RT cause Alzheimer's disease."; Ann. Neurol. 44:965-967(1998).
 RL [18]
 RN
 RP VARIANT GLY-318.
 RX MEDLINE=99066775; PubMed=9851450;
 RA Aldudo J., Bullido M.J., Frank A., Valdivieso F.;
 RT "Missense mutation E318G of the presenilin-1 gene appears to be a
 RT nonpathogenic polymorphism."; Ann. Neurol. 44:985-986(1998).
 RL [19]
 RN
 RP VARIANTS AD VAL-79; CYS-115; VAL-231, AND VARIANT GLY-318.
 RX MEDLINE=98046005; PubMed=9384602;
 RA Cruts M., van Duijn C.M., Backhovens H., van den Broeck M.,
 RA Wehnert A., Serneels S., Sherrington R., Hutton M., Hardy J.,
 RA St George-Hyslop P.H., Hofman A., van Broeckhoven C.;
 RT "Estimation of the genetic contribution of presenilin-1 and -2
 RT mutations in a population-based study of presenile Alzheimer
 RT disease."; Hum. Mol. Genet. 7:43-51(1998).
 RL [20]
 RN
 RP VARIANTS AD D-120; R-163; V-209; V-260; L-264; Y-410 AND P-426.
 RX MEDLINE=98180720; PubMed=9521423;
 RA Poorkaj P., Sharma V., Anderson L., Nemens E., Alonso M.E., Orr H.,
 RA White J., Heston L., Bird T.D., Schellenberg G.B.;
 RT "Missense mutations in the chromosome 14 familial Alzheimer's disease
 RT presenilin 1 gene."; Hum. Mutat. 11:216-221(1998).
 RL [21]
 RN
 RP VARIANT AD GLU-378.
 RX MEDLINE=99211215; PubMed=10200054;
 RA Besancon R., Lorenzi A., Cruts M., Radawiec S., Sturtz F.,

Query Match 90.0%; Score 36; DB 1; Length 467;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VNMAEGD 8
DB 296 VNMAEGD 302

RESULT 2
PSN1_MICMU STANDARD; PRT; 467 AA.

ID PSN1_MICMU
AC P79802;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PRESENILIN 1 (PS-1)
GN PSN1 OR PSN1 OR PSI.
OS Microcebus murinus (Lesser mouse lemur).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Strepsirhini; Cheirogaleidae;
OC Microcebus.
OX NCBI_TaxID=30608;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Calenda A., Mestre-Frances N., Czech C., Pradier L., Bons N.,
RA Bellis M.;
RT "Molecular cloning, sequencing, and brain expression of the
presenilin 1 gene in Microcebus murinus.";
RL Biochem. Biophys. Res. Commun. 228:430-439(1996).
CC -1- FUNCTION: MAY PLAY A ROLE IN INTRACELLULAR SIGNALING AND GENE
EXPRESSION OR IN LINKING CHROMATIN TO THE NUCLEAR MEMBRANE.
CC MAY FUNCTION IN THE CYTOPLASMIC PARTITIONING OF PROTEINS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. GOLGI AND
ENDOPLASMIC RETICULUM (BY SIMILARITY).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; I-467 (SHOWN HERE) AND I-
463; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: FOUND PREDOMINANTLY IN NEURONS OF THE
DIFFERENT CORTICAL LAYERS AND HIPPOCAMPUS BUT ALSO IN SUBCORTICAL
STRUCTURES.
CC -1- SIMILARITY: BELONGS TO THE PRESENILIN FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; 271333; CAA95930.1; -;
DR MEROPS; A22.001; -;
DR InterPro; IPR001108; -;
DR Pfam; PF01080; Presenilin; 1.
DR PRINTS; PR01072; PRESENILIN.
DR PRINTS; PR01073; PRESENILIN.
KW Transmembrane; Glycoprotein; Endoplasmic reticulum; Golgi stack;
KW Alternative splicing.
FT TRANSMEM 83 103 POTENTIAL.
FT TRANSMEM 133 153 POTENTIAL.
FT TRANSMEM 161 181 POTENTIAL.
FT TRANSMEM 191 211 POTENTIAL.
FT TRANSMEM 221 241 POTENTIAL.
FT TRANSMEM 244 264 POTENTIAL.
FT TRANSMEM 281 301 POTENTIAL.
FT TRANSMEM 408 428 POTENTIAL.
FT TRANSMEM 433 453 POTENTIAL.
FT CARBOHYD 279 293 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 405 409 N-LINKED (GLCNAC. .) (POTENTIAL).

FT VARSPLIC 26 29 MISSING (IN ISOFORM I-463).
SQ SEQUENCE 467 AA; 52384 MW; D986FF2CA7F2975C CRC64;

Query Match 90.0%; Score 36; DB 1; Length 467;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VNMAEGD 8
DB 296 VNMAEGD 302

RESULT 3
PSN1_MOUSE STANDARD; PRT; 467 AA.

ID PSN1_MOUSE
AC P49769;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PRESENILIN 1 (PS-1) (S182 PROTEIN).
GN PSN1 OR PSN1 OR AD3H.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=95319502; PubMed=7596406;
RA Sherrington R., Rogeev E.I., Liang Y., Rogeev E.A., Levesque G.,
RA Ikeda M., Chi H., Lin C., Li G., Holman K., Tsuda T., Mar L.,
RA Foncin J.-F., Bruni A.C., Montesi M.P., Sorbi S., Rainero I.,
RA Pinessi L., Nee L., Chumakov I., Pollen D., Brookes A.,
RA Sausse P., Polinsky R.J., Wasco W., da Silva H.A.R., Haines J.L.,
RA Pericak-Vance M.A., Tanzi R.E., Roses A.D., Fraser P.E.,
RA Rommens J.M., St George-Hyslop P.H.;
RT "Cloning of a gene bearing missense mutations in early-onset familial
Alzheimer's disease.";
RL Nature 375:754-760(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SVJ;
RX MEDLINE=97442406; PubMed=9295283;
RA Mitsuda N., Roses A.D., Vitek M.P.;
RT "Transcriptional regulation of the mouse presenilin-1 gene.";
RL J. Biol. Chem. 272:23489-23497(1997).
CC -1- FUNCTION: MAY PLAY A ROLE IN INTRACELLULAR SIGNALING AND GENE
EXPRESSION OR IN LINKING CHROMATIN TO THE NUCLEAR MEMBRANE.
CC MAY FUNCTION IN THE CYTOPLASMIC PARTITIONING OF PROTEINS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. GOLGI AND
ENDOPLASMIC RETICULUM (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE PRESENILIN FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; L42177; AAC42094.1; -;
DR EMBL; AF007560; AAB72049.1; -;
DR MEROPS; A22.001; -;
DR MGD; MGI:1202717; Psen1.
DR InterPro; IPR001108; -;
DR Pfam; PF01030; Presenilin; 1.
DR PRINTS; PR01072; PRESENILIN.
DR PRINTS; PR01073; PRESENILIN.
KW Transmembrane; Glycoprotein; Endoplasmic reticulum; Golgi stack.
KW TRANSMEM 83 103 POTENTIAL.

FT	TRANSMEM	133	153	POTENTIAL.	DR	Pfam; PF01080; Presenilin; 1.
FT	TRANSMEM	161	181	POTENTIAL.	DR	PRINTS; PR01072; PRESENILIN.
FT	TRANSMEM	195	215	POTENTIAL.	DR	PRINTS; PR01073; PRESENILIN.
FT	TRANSMEM	221	241	POTENTIAL.	KW	Transmembrane; Glycoprotein; Endoplasmic reticulum; Golgi stack.
FT	TRANSMEM	244	264	POTENTIAL.	FT	TRANSMEM
FT	TRANSMEM	281	301	POTENTIAL.	FT	TRANSMEM
FT	TRANSMEM	408	428	POTENTIAL.	FT	TRANSMEM
FT	TRANSMEM	433	453	POTENTIAL.	FT	TRANSMEM
FT	CARBOHYD	30	30	N-LINKED (GLCNAC. . .) (POTENTIAL).	FT	TRANSMEM
FT	CARBOHYD	279	279	N-LINKED (GLCNAC. . .) (POTENTIAL).	FT	TRANSMEM
FT	CARBOHYD	405	405	N-LINKED (GLCNAC. . .) (POTENTIAL).	FT	TRANSMEM
SQ	SEQUENCE	467 AA;	52639 MW;	D07215B4BAD2D549 CRC64;	FT	TRANSMEM
QY	2 VNMAEGD 8				FT	TRANSMEM
DB	296 VNMAEGD 302				FT	TRANSMEM
RESULT	4				FT	TRANSMEM
ID	PSN1_RAT	STANDARD;	PRT;	468 AA.	FT	TRANSMEM
AC	P97887; p97529;				FT	TRANSMEM
DT	15-JUL-1999 (Rel. 38, Created)				FT	TRANSMEM
DT	15-JUL-1999 (Rel. 38, Last sequence update)				FT	TRANSMEM
DT	01-OCT-2000 (Rel. 40, Last annotation update)				FT	TRANSMEM
DE	PRESENILIN 1 (PS-1) (S182 PROTEIN).				FT	TRANSMEM
GN	PSN1 OR PSNLI				FT	TRANSMEM
OS	Rattus norvegicus (Rat).				FT	TRANSMEM
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				FT	TRANSMEM
OC	Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Rattus.				FT	TRANSMEM
OX	NCBI_TaxID=10116;				FT	TRANSMEM
RN	[1]				FT	TRANSMEM
RP	SEQUENCE FROM N.A.				FT	TRANSMEM
RC	STRAIN-WISTAR; TISSUE-Brain;				FT	TRANSMEM
RX	MEDLINE=97199371; PubMed=9047347;				FT	TRANSMEM
RA	Taniguchi T., Hashimoto T., Taniguchi R., Shimada K., Kawamata T.,				FT	TRANSMEM
RA	Yasuda M., Nakai M., Terashima A., Koizumi T., Maeda K., Tanaka C.;				FT	TRANSMEM
RT	"Cloning of the cDNA encoding rat presenilin-1.";				FT	TRANSMEM
RL	Gene 186:73-75(1997).				FT	TRANSMEM
RN	[2]				FT	TRANSMEM
RP	SEQUENCE FROM N.A.				FT	TRANSMEM
RC	STRAIN-WISTAR; TISSUE-Brain;				FT	TRANSMEM
RX	MEDLINE=96255262; PubMed=8710164;				FT	TRANSMEM
RA	Takahashi H., Murayama M., Takashima A., Mercken M., Nakazato Y.,				FT	TRANSMEM
RA	Noguchi K., Imahori K.;				FT	TRANSMEM
RT	"Molecular cloning and expression of the rat homologue of				FT	TRANSMEM
RT	presenilin-1.";				FT	TRANSMEM
RL	Neurosci. Lett. 206:113-116(1996).				FT	TRANSMEM
CC	-1- FUNCTION: MAY PLAY A ROLE IN INTRACELLULAR SIGNALING AND GENE				FT	TRANSMEM
CC	EXPRESSION OR IN LINKING CHROMATIN TO THE NUCLEAR MEMBRANE.				FT	TRANSMEM
CC	MAY FUNCTION IN THE CYTOPLASMIC PARTITIONING OF PROTEINS.				FT	TRANSMEM
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. GOLGI AND				FT	TRANSMEM
CC	ENDOPLASMIC RETICULUM (BY SIMILARITY).				FT	TRANSMEM
CC	-1- SIMILARITY: BELONGS TO THE PRESENILIN FAMILY.				FT	TRANSMEM
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				FT	TRANSMEM
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation				FT	TRANSMEM
CC	the European Bioinformatics Institute. There are no restrictions on its				FT	TRANSMEM
CC	use by non-profit institutions as long as its content is in no way				FT	TRANSMEM
CC	modified and this statement is not removed. Usage by and for commercial				FT	TRANSMEM
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				FT	TRANSMEM
CC	or send an email to license@isb-sib.ch).				FT	TRANSMEM
CC	-----				FT	TRANSMEM
CC	EMBL; D82578; BAAL1575.1; .				FT	TRANSMEM
DR	EMBL; D82363; BAAL1564.1; .				FT	TRANSMEM
DR	MEROPS; A22.001; .				FT	TRANSMEM
DR	InterPro; IPR001108; .				FT	TRANSMEM
DR	InterPro; IPR002031; .				FT	TRANSMEM
QY	2 VNMAEGD 8				FT	TRANSMEM
DB	296 VNMAEGD 302				FT	TRANSMEM
Query Match	90.0%;	Score 36;	DB 1;	Length 467;	FT	TRANSMEM
Best Local Similarity	100.0%;	Pred. No. 1.2;			FT	TRANSMEM
Matches	7;	Conservative	0;	Mismatches	0;	Gaps
QY	2 VNMAEGD 8				FT	TRANSMEM
DB	296 VNMAEGD 302				FT	TRANSMEM
Query Match	90.0%;	Score 36;	DB 1;	Length 468;	FT	TRANSMEM
Best Local Similarity	100.0%;	Pred. No. 1.2;			FT	TRANSMEM
Matches	7;	Conservative	0;	Mismatches	0;	Indels
QY	2 VNMAEGD 8				FT	TRANSMEM
DB	296 VNMAEGD 302				FT	TRANSMEM
RESULT	5				FT	TRANSMEM
ID	PSN1_BOVIN	STANDARD;	PRT;	478 AA.	FT	TRANSMEM
AC	Q9XT97;				FT	TRANSMEM
DT	01-OCT-2000 (Rel. 40, Created)				FT	TRANSMEM
DT	01-OCT-2000 (Rel. 40, Last sequence update)				FT	TRANSMEM
DT	01-OCT-2000 (Rel. 40, Last annotation update)				FT	TRANSMEM
DE	PRESENILIN 1 (PS-1).				FT	TRANSMEM
GN	PSN1.				FT	TRANSMEM
OS	Bos taurus (Bovine).				FT	TRANSMEM
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				FT	TRANSMEM
OC	Mammalia; Euthera; Cetartiodactyla; Ruminantia; Pecora; Bovidae;				FT	TRANSMEM
OC	Bovidae; Bovinae; Bos.				FT	TRANSMEM
OX	NCBI_TaxID=9913;				FT	TRANSMEM
RN	[1]				FT	TRANSMEM
RP	SEQUENCE FROM N.A.				FT	TRANSMEM
RC	TISSUE-Brain;				FT	TRANSMEM
RA	Sahara N., Shirasawa T., Mori H.;				FT	TRANSMEM
RT	"Molecular cloning of bovine presenilin 1 gene.";				FT	TRANSMEM
RL	Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.				FT	TRANSMEM
CC	-1- FUNCTION: MAY PLAY A ROLE IN INTRACELLULAR SIGNALING AND GENE				FT	TRANSMEM
CC	EXPRESSION OR IN LINKING CHROMATIN TO THE NUCLEAR MEMBRANE. MAY				FT	TRANSMEM
CC	FUNCTION IN THE CYTOPLASMIC PARTITIONING OF PROTEINS (BY				FT	TRANSMEM
CC	SIMILARITY).				FT	TRANSMEM
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. GOLGI AND				FT	TRANSMEM
CC	ENDOPLASMIC RETICULUM (BY SIMILARITY).				FT	TRANSMEM
CC	-1- SIMILARITY: BELONGS TO THE PRESENILIN FAMILY.				FT	TRANSMEM
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				FT	TRANSMEM
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation				FT	TRANSMEM
CC	the European Bioinformatics Institute. There are no restrictions on its				FT	TRANSMEM
CC	use by non-profit institutions as long as its content is in no way				FT	TRANSMEM
CC	modified and this statement is not removed. Usage by and for commercial				FT	TRANSMEM
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				FT	TRANSMEM
CC	or send an email to license@isb-sib.ch).				FT	TRANSMEM
CC	-----				FT	TRANSMEM
CC	EMBL; AF038936; AAD39023.1; .				FT	TRANSMEM
DR	InterPro; IPR001108; .				FT	TRANSMEM
DR	InterPro; IPR002031; .				FT	TRANSMEM
DR	Pfam; PF01080; Presenilin; 1.				FT	TRANSMEM
DR	PRINTS; PR01072; PRESENILIN.				FT	TRANSMEM
DR	PRINTS; PR01073; PRESENILIN.				FT	TRANSMEM
KW	Transmembrane; Glycoprotein; Endoplasmic reticulum; Golgi stack.				FT	TRANSMEM

```
FT TRANSNM 83 103 POTENTIAL.
FT TRANSNM 134 154 POTENTIAL.
FT TRANSNM 162 182 POTENTIAL.
FT TRANSNM 196 216 POTENTIAL.
FT TRANSNM 222 242 POTENTIAL.
FT TRANSNM 245 265 POTENTIAL.
FT TRANSNM 282 302 POTENTIAL.
FT TRANSNM 392 412 POTENTIAL.
FT TRANSNM 419 439 POTENTIAL.
FT TRANSNM 444 464 POTENTIAL.
FT CARBOHYD 30 30 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 280 280 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 416 416 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 478 AA: 53653 MW; 59E3FC0A1010D906 CRC64;

Query Match 90.0%; Score 36; DB 1; Length 478;
Best Local Similarity 100.0%; Pred. No. 1.2; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

QY 2 VNNAEGD 8
DB 297 VNNAEGD 303

RESULT 6
PSNL_XENLA STANDARD; PRT; 433 AA.
AC O12976;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PRESENTILIN ALPHA.
GN PS-ALPHA.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97223465; PubMed=9070286;
RA Tsujimura A., Yasojima K., Hashimoto-Gotoh T.:
RT "Cloning of Xenopus presentilin-alpha and -beta cDNAs and their
RT differential expression in oogenesis and embryogenesis.";
RL Biochem. Biophys. Res. Commun. 231:392-396(1997).
CC -!- FUNCTION: MAY PLAY A ROLE IN NEGATIVE REGULATION OF APOPTOTIC
CC CASCADES DURING OOGENESIS AND EMBRYOGENESIS, AND IN
CC DEVELOPMENTALLY MATURED TISSUES SUCH AS BRAIN TISSUE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN OVARIES AND TO A LESSER
CC EXTENT IN TESTIS, INTESTINE, KIDNEY, BRAIN, EYE AND LUNG. WEAK
CC EXPRESSION IN LIVER AND HEART. PRESENT IN TRACE AMOUNTS IN
CC SKELETAL MUSCLE.
CC -!- DEVELOPMENTAL STAGE: ABUNDANT IN EARLY STAGES OF OOGENESIS. THE
CC EXPRESSION IS RAPIDLY REDUCED BETWEEN MEIOTIC MATURATION AND
CC FERTILIZATION STAGES.
CC -!- SIMILARITY: BELONGS TO THE PRESENTILIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D84427; BAAL19570.1;
CC DR MEROPS; A22.001;
CC DR InterPro; IPR001108;
CC DR Pfam; PF01080; Presentilin; 1.
CC DR PRINTS; PR01072; PRESENTILIN.
```

```
KW Transmembrane; Glycoprotein.
FT TRANSNM 48 68 POTENTIAL.
FT TRANSNM 99 119 POTENTIAL.
FT TRANSNM 127 147 POTENTIAL.
FT TRANSNM 161 181 POTENTIAL.
FT TRANSNM 184 204 POTENTIAL.
FT TRANSNM 210 230 POTENTIAL.
FT TRANSNM 247 267 POTENTIAL.
FT TRANSNM 374 394 POTENTIAL.
FT TRANSNM 399 419 POTENTIAL.
FT CARBOHYD 245 245 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 371 371 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 433 AA: 48301 MW; 71CCB3F6BB9C0AF CRC64;

Query Match 82.5%; Score 33; DB 1; Length 433;
Best Local Similarity 85.7%; Pred. No. 5.3;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 VNNAEGD 8
DB 262 VNNAEGD 268

RESULT 7
GR75_CRIGR STANDARD; PRT; 679 AA.
AC O35501;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE MITOCHONDRIAL STRESS-70 PROTEIN PRECURSOR (75 KDA GLUCOSE REGULATED
DE PROTEIN) (GRP 75).
GN HSPA9.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
OX NCBI_TaxID=10029;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97405843; PubMed=9260887;
RA Singh B., Soltyz B.J., Wu Z.C., Patel H.V., Freeman K.B., Gupta R.S.;
RT "Cloning and some novel characteristics of mitochondrial Hsp70 from
RT Chinese hamster cells.";
RL Exp. Cell Res. 234:205-216(1997).
CC -!- FUNCTION: IMPLICATED IN THE CONTROL OF CELL PROLIFERATION AND
CC CELLULAR AGING. MAY ALSO ACT AS A CHAPERONE (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U92313; AAB62091.1;
CC DR HSP; P04475; IDK2.
CC DR InterPro; IPR001023;
CC DR Pfam; PF00012; HSP70; 1.
CC DR PRINTS; PR00301; HEATSHOCK70.
CC DR PROSITE; PS00297; HSP70_1; 1.
CC DR PROSITE; PS00329; HSP70_2; 1.
CC DR PROSITE; PS01036; HSP70_3; 1.
CC DR ATP-binding; Mitochondrion; Transit peptide.
CC
CC CHAIN 1 46 MITOCHONDRIAL (BY SIMILARITY).
CC FT TRANSIT 47 679 MITOCHONDRIAL STRESS-70 PROTEIN.
CC SQ SEQUENCE 679 AA; 73730 MW; CC652241B12ECA70 CRC64;
```

```
Query Match 77.5%; Score 31; DB 1; Length 679;
Best Local Similarity 85.7%; Pred. No. 25;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVNMAEG 7
Db 581 AVNMAEG 587

RESULT 8
GR75_HUMAN
ID GR75_HUMAN STANDARD; PRT; 679 AA.
AC P38646; P30036; P31932;
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE MITOCHONDRIAL STRESS-70 PROTEIN PRECURSOR (75 KDA GLUCOSE REGULATED
DE PROTEIN) (GRP 75) (PEPTIDE-BINDING PROTEIN 74) (PBP74) (MORTALIN)
DE (MOT).
GN HSPA9B OR HSPA9 OR GRP75.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95130547; PubMed=7829505;
RA Bhattacharyya T., Karnesis A.N., Murphy S.P., Hoang T., Freeman B.C.,
RA Phillips B., Morimoto R.I.;
RT "Cloning and subcellular localization of human mitochondrial hsp70.";
RL J. Biol. Chem. 270:1705-1710(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=B-cell;
RX MEDLINE=93268309; PubMed=7684501;
RA Domancic S.Z., Denagel D.C., Dahlseid J.N., Green J.M.;
RA Pierce S.K.;
RT "Cloning of the gene encoding peptide-binding protein 74 shows that
RT it is a new member of the heat shock protein 70 family.";
RL Mol. Cell. Biol. 13:3598-3610(1993).
RN [3]
RP SEQUENCE OF 47-56.
RC TISSUE=Liver;
RX MEDLINE=93162045; PubMed=1286669;
RA Hochstrasser D.F., Frutiger S., Paquet N., Bairoch A., Ravier F.,
RA Pasquali C., Sanchez J.-C., Tissot J.-D., Bjellqvist B., Vargas R.,
RA Appel R.D., Hughes G.J.;
RT "Human liver protein map: a reference database established by
RT microsequencing and gel comparison.";
RL Electrophoresis 13:992-1001(1992).
RN [4]
RP REVISIONS.
RC TISSUE=Liver;
RX MEDLINE=94147969; PubMed=8313870;
RA Hughes G.J., Frutiger S., Paquet N., Pasquali C., Sanchez J.-C.,
RA Tissot J.-D., Bairoch A., Appel R.D., Hochstrasser D.F.;
RT "Human liver protein map: update 1993.";
RL Electrophoresis 14:1216-1222(1993).
RN [5]
RP SEQUENCE OF 47-68.
RC TISSUE=Colon carcinoma;
RX MEDLINE=97295306; PubMed=9150948;
RA Ji H., Reid G.E., Moritz R.L., Edges J.S., Burgess A.W., Simpson R.J.;
RA Tisot J.-D., Bairoch A., Appel R.D., Hochstrasser D.F.;
RT "A two-dimensional gel database of human colon carcinoma proteins.";
RL Electrophoresis 18:605-613(1997).
CC -1- FUNCTION: IMPLICATED IN THE CONTROL OF CELL PROLIFERATION AND
CC CELLULAR AGING. MAY ALSO ACT AS A CHAPERONE.
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; L15189; AAA67526.1;
CC EMBL; L11066; -; NOT_ANNOTATED_CDS.
CC PIR; B48127; B48127.
CC HSP; P04475; IDKX.
CC SWISS-2DPAGE; P38646; HUMAN.
CC HSC-2DPAGE; P38646; HUMAN.
CC MIM; 600548;
CC InterPro: IPR001023;
CC Pfam: PF00012; HSP70; 1.
CC PRINTS; PR00301; HEATSHOCK70.
CC PROSITE; PS00297; HSP70_1; 1.
CC PROSITE; PS00329; HSP70_2; 1.
CC PROSITE; PS01036; HSP70_3; 1.
CC ATP-binding; Mitochondrion; Transit peptide.
KW TRANSIT 1 46 MITOCHONDRIAL.
FT CHAIN 47 679 MITOCHONDRIAL STRESS-70 PROTEIN.
FT CONFLICT 540 540 R -> G (IN REF. 2).
FT SEQUENCE 679 AA; 73779 MW; 908D9B3606D57653 CRC64;
SQ
Query Match 77.5%; Score 31; DB 1; Length 679;
Best Local Similarity 85.7%; Pred. No. 25;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVNMAEG 7
Db 581 AVNMAEG 587

RESULT 9
GR75_MOUSE
ID GR75_MOUSE STANDARD; PRT; 679 AA.
AC P38647;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE MITOCHONDRIAL STRESS-70 PROTEIN PRECURSOR (75 KDA GLUCOSE REGULATED
DE PROTEIN) (GRP 75) (PEPTIDE-BINDING PROTEIN 74) (PBP74) (P66 MOT)
DE (MORTALIN).
GN HSPA9A OR HSPA9 OR GRP75 OR HSP74.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CD-1-ICR; TISSUE=Embryonic fibroblast;
RX MEDLINE=93203261; PubMed=8454632;
RA Wadhwa R., Kaul S.C., Ikawa Y., Sugimoto Y.;
RT "Identification of a novel member of mouse hsp70 family. Its
RT association with cellular mortal phenotype.";
RL J. Biol. Chem. 268:6615-6621(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CD-1-ICR; TISSUE=Embryonic fibroblast;
RX MEDLINE=94042962; PubMed=7693662;
RA Wadhwa R., Kaul S.C., Sugimoto Y., Mitsui Y.;
RT "Induction of cellular senescence by transfection of cytosolic
RT mortalin cDNA in NIH 3T3 cells.";
RL J. Biol. Chem. 268:22239-22242(1993).
RN [3]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=B-cell;
RX MEDLINE=93268309; PubMed=7684501;
RA Domancic S.Z., Denagel D.C., Dahlseid J.N., Green J.M., Pierce S.K.;
RT "Cloning of the gene encoding peptide-binding protein 74 shows that
RT it is a new member of the heat shock protein 70 family.";
RL Mol. Cell. Biol. 13:3598-3610(1993).
```


DR PROSITE; PS00329; HSP70_2; 1.
 DR PROSITE; PS01036; HSP70_3; 1.
 KW ATP-binding; Mitochondrion; Transit peptide.
 FT CHAIN 1 46 MITOCHONDRION (BY SIMILARITY).
 FT TRANSIT 47 679 MITOCHONDRIAL STRESS-70 PROTEIN.
 FT CHAIN 37 37 V -> A (IN REF. 2).
 FT CONFLICT 81 81 S -> A (IN REF. 2).
 FT CONFLICT 373 373 R -> A (IN REF. 2).
 FT CONFLICT 589 589 V -> I (IN REF. 2).
 SQ SEQUENCE 679 AA; 73859 MW; 85405F60EF1889C9 CRC64;

Query Match 77.5%; Score 31; DB 1; Length 679;
 Best Local Similarity 85.7%; Pred. No. 25;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVNMAEG 7
 Db 581 AVNMAEG 587
 :|||||

RESULT 11
 GP42_RAT STANDARD; PRT; 233 AA.
 AC P23505;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 01-NOV-1991 (Rel. 20, Last annotation update)
 DE CELL SURFACE GLYCOPROTEIN GP42 PRECURSOR.
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91086851; PubMed=1845873;
 RA Seaman W.E., Niemi E.C., Stark M.R., Goldfien R.D., Pollock A.S.,
 RA Imboden J.B.;
 RT "Molecular cloning of gp42, a cell-surface molecule that is
 RT selectively induced on rat natural killer cells by interleukin 2:
 RT glycolipid membrane anchoring and capacity for transmembrane
 RT signaling.";
 RL J. Exp. Med. 173:251-260(1991).
 CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
 CC -!- TISSUE SPECIFICITY: NK CELLS.
 CC -!- INDUCTION: BY IL-2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X56448; CAA39831.1; -;
 DR PIR; JH0372; JH0372.
 DR InterPro; IPR003006; -;
 DR Pfam; PF00047; Ig; 2.
 KW Signal; Glycoprotein; GPI-anchor; Membrane.
 FT SIGNAL 1 16
 FT CHAIN 17 217 CELL SURFACE GLYCOPROTEIN GP42.
 FT PROPEP 218 233 REMOVED IN MATURE FORM (POTENTIAL).
 FT LIPID 217 217 GPI-ANCHOR (POTENTIAL).
 FT CARBOHYD 29 29 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 66 66 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 215 215 V -> M (IN CLONE 2).
 SQ SEQUENCE 233 AA; 26025 MW; E6A03816FE151C59 CRC64;

Query Match 75.0%; Score 30; DB 1; Length 233;
 Best Local Similarity 62.5%; Pred. No. 13;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SVNMAEG 8
 Db 123 ATNLAEGD 130
 :|||||

RESULT 12
 MURF_MYCTU STANDARD; PRT; 510 AA.
 AC O06220;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE UDP-N-ACETYLURAMUOYL-L-ALANYL-2,6-DIAMINOPIMELATE--D-ALANYL-D-
 DE ALANYL LIGASE (EC 6.3.2.15) (UDP-MURNAC-PENTAPEPTIDE SYNTHETASE)
 DE (D-ALANYL-D-ALANINE-ADDING ENZYME).
 GN MURF OR RV2157C OR MTCY270.11.
 OS Mycobacterium tuberculosis.
 CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 CC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
 RA Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 CC -!- FUNCTION: INVOLVED IN CELL WALL FORMATION. CATALYSES THE FINAL
 CC STEP IN THE SYNTHESIS OF UDP-N-ACETYLMURAMOYL-PENTAPEPTIDE. THE
 CC PRECURSOR OF MUREIN (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: ATP + UDP-N-ACETYLMURAMOYL-L-ALANYL-D-GLUTAMYL
 CC -MESO-2,6-DIAMINOHEPTANEDIOATE + D-ALANYL-D-ALANYL - ADP +
 CC ORTHOPHOSPHATE + UDP-N-ACETYLMURAMOYL-L-ALANYL-D-GLUTAMYL-6-
 CC CARBOXY-L-LYSYL-D-ALANYL-D-ALANINE.
 CC -!- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.
 CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
 CC -!- SIMILARITY: BELONGS TO THE MURCDEF FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z95388; CAB08670.1; -;
 DR TuberculList; Rv2157c; -;
 DR InterPro; IPR000713; -;
 DR Pfam; PF01225; Mur_ligase; 1.
 KW Peptidoglycan synthesis; Cell division; Cell wall; Ligase;
 FT NP_BIND 136 142 ATP (POTENTIAL).
 SQ SEQUENCE 510 AA; 51632 MW; 7BE83EDD1BAA2EC4 CRC64;

Query Match 75.0%; Score 30; DB 1; Length 510;
 Best Local Similarity 62.5%; Pred. No. 31;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVNMAEG 8
 :|||||

Db 459 AVNAVGD 466

RESULT 13

GSPD_VIBCH STANDARD: PRT: 674 AA.
AC P45779;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE GENERAL SECRETION PATHWAY PROTEIN D PRECURSOR (CHOLERA TOXIN SECRETION
DE PROTEIN EPSPD)
GN EPSPD OR VC2733.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR TRH7000;
RA Overbye L.J.;
RT "Organization of the general secretion pathway genes in Vibrio
RT cholerae";
RL Thesis (1994), Michigan State University / East Lansing, U.S.A.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae";
RL Nature 406:477-483(2000).
CC -!- FUNCTION: REQUIRED FOR SECRETION OF CHOLERA TOXIN THROUGH THE
CC OUTER MEMBRANE.
CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE EXED/OUTD/PULD/XPSD FAMILY.
CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT (POSSIBLY NATURAL) IN POSITION 55.
CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; L33796; AAA58785.1;
DR EMBL; AE004338; ; NOT_ANNOTATED_CDS.
DR TIGR; VC2733;
DR InterPro; IPR000016; ;
DR InterPro; IPR001775; ;
DR Pfam; PF00263; Bac_GSPproteins; 1.
DR PRINTS; PR00811; BCTERIALGSPD.
DR PROSITE; PS00875; T2SP_D; 1.
KW Transport; Outer membrane; Signal.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 674 GENERAL SECRETION PATHWAY PROTEIN D.
FT CONFLICT 89 89 V -> A (IN REF. 1).
FT CONFLICT 144 144 R -> P (IN REF. 1).
SQ SEQUENCE 674 AA; 73469 MW; 3D77B891A59E6223 CRC64;

Query Match 75.0%; Score 30; DB 1; Length 674;
Best Local Similarity 85.7%; Pred. No. 42;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 VNNAEGD 8
RN [1]

Db 351 VEMAECD 357

RESULT 14

GYRA_MYCSM STANDARD: PRT: 842 AA.
AC P48354;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE DNA GYRASE SUBUNIT A (EC 5.99.1.3).
GN GYRA.
OS Mycobacterium smegmatis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1772;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 700084 / MC(2)155;
RX MEDLINE=97032832; PubMed=8878580;
RA Revel-Viravau V., Truong Q.C., Moreau N., Jarlier V., Sougakoff W.;
RT "Sequence analysis, purification, and study of inhibition by 4-
RT quinolones of the DNA gyrase from Mycobacterium smegmatis";
RL Antimicrob. Agents Chemother. 40:2054-2061(1996).
CC -!- FUNCTION: DNA GYRASE NEGATIVELY SUPERCOILS CLOSED CIRCULAR DOUBLE-
CC STRANDED DNA IN AN ATP-DEPENDENT MANNER AND ALSO CATALYZES THE
CC INTERCONVERSION OF OTHER TOPOLOGICAL ISOMERS OF DOUBLE-STRANDED
CC DNA RINGS, INCLUDING CATENANES AND KNOTTED RINGS.
CC -!- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
CC OF DOUBLE-STRANDED DNA.
CC -!- SUBUNIT: MADE UP OF TWO CHAINS. THE A CHAIN IS RESPONSIBLE FOR DNA
CC BREAKAGE AND REJOINING; THE B CHAIN CATALYZES ATP HYDROLYSIS. THE
CC ENZYME FORMS AN A2B2 TETRAMER.
CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; X94224; CAA63918.1;
DR InterPro; IPR002205; ;
DR Pfam; PF00521; DNA_topoisoIV; 1.
KW Topoisomerase; Isomerase; ATP-binding; Antibiotic resistance.
FT ACT_SITE 130 130 DNA CLEAVAGE (BY SIMILARITY).
SQ SEQUENCE 842 AA; 93183 MW; 50AE5DBE80ED45DD CRC64;

Query Match 75.0%; Score 30; DB 1; Length 842;
Best Local Similarity 71.4%; Pred. No. 53;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 VNNAEGD 8
RN [1]

RESULT 15

YBGF_PSEPU STANDARD: PRT: 268 AA.
ID YBGF_PSEPU
AC P43037;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 28.7 KDa PROTEIN IN PAL 5'REGION PRECURSOR (ORF2).
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=303;
RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=MT-2;
 RX MEDLINE=96198174; PubMed=8626299;
 RA Rodriguez-Herva J.J., Ramos-Gonzalez M.I., Ramos J.L.;
 RT "The Pseudomonas putida peptidoglycan-associated outer membrane
 lipoprotein is involved in maintenance of the integrity of the cell
 cell envelope.";
 RL J. Bacteriol. 178:1699-1706(1996).
 CC -!- SIMILARITY: TO E.COLI YBGF.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X74218; CAA52295.1; -;
 KW Hypothetical protein; Signal.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 / 268 HYPOTHETICAL 28.7 KDA PROTEIN IN PAL
 FT 5' REGION.
 SQ SEQUENCE 268 AA; 28730 MW; 3183F1F06B1E2A21 CRC64;

Query Match 72.5%; Score 29; DB 1; Length 268;
 Best Local Similarity 71.4%; Pred. No. 26;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 VNMAGD 8
 DB 194 VNLAKGD 200

Search completed: September 6, 2001, 16:51:06
 Job time: 810 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:49:46 ; Search time 231.42 Seconds
(without alignments)
4.574 Million cell updates/sec

Title: US-09-603-713-6
Perfect score: 40
Sequence: 1 SVNMAEGD 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_16:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_unclassified:*
- 13: sp_vertebrate:*
- 14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	ID	Description
1	36	90.0	469	11	Q9JLP9	Q9jlp9 mus musculu
2	34	85.0	562	2	P71654	P71654 mycobacteri
3	32	80.0	71	2	Q9XCD6	Q9xcd6 mycobacteri
4	32	80.0	292	5	Q27109	Q27109 trichomonas
5	32	80.0	354	3	O60121	O60121 schizosacch
6	32	80.0	449	10	Q9SMC2	Q9smc2 nicotiana p
7	32	80.0	797	2	Q9S341	Q9s341 photorhabdu
8	32	80.0	1064	10	Q9MAU0	Q9mau0 arabidopsi
9	31	77.5	55	5	O44360	O44360 ceratitis c
10	31	77.5	205	4	Q9H018	Q9h018 homo sapien
11	31	77.5	282	4	Q9NVT9	Q9nvt9 homo sapien
12	31	77.5	300	1	Q9HML4	Q9hml4 halobacteri
13	31	77.5	377	2	Q9F6W5	Q9f6w5 chloroflexu
14	31	77.5	403	10	Q9LLP3	Q9llp3 oryza sativ
15	31	77.5	559	4	Q9P0W8	Q9p0w8 homo sapien
16	31	77.5	567	4	Q9H0X0	Q9h0x0 homo sapien
17	31	77.5	849	2	Q9KN18	Q9kni8 vibrio chol
18	31	77.5	1750	5	O96610	O96610 drosophila
19	31	77.5	1993	5	Q9VGH9	Q9vgh9 drosophila

20	30	75.0	90	13	Q92083	Q92083 makaira nig
21	30	75.0	90	13	Q92121	Q92121 xiphias gla
22	30	75.0	181	10	Q9M718	Q9m718 perseia amer
23	30	75.0	219	10	Q9FQ28	Q9fq28 nicotiana t
24	30	75.0	331	13	O42112	O42112 brachydanio
25	30	75.0	334	5	Q9XVP9	Q9xvp9 caenorhabdi
26	30	75.0	386	10	O23305	O23305 arabidopsi
27	30	75.0	488	8	Q9XPY1	Q9xpy1 hecatonema
28	30	75.0	724	14	Q9Q3G5	Q9q3g5 turkey astr
29	30	75.0	854	2	O59556	O59556 mycobacteri
30	30	75.0	920	5	O9GRN4	O9grn4 leishmania
31	30	75.0	930	11	Q9ET63	Q9et63 mus musculu
32	30	75.0	930	11	Q9EQR2	Q9eqr2 mus musculu
33	30	75.0	1022	6	Q28628	Q28628 oryctolagus
34	29	72.5	97	2	P72476	P72476 streptococc
35	29	72.5	152	5	O26991	O26991 tritrichomo
36	29	72.5	274	2	O9RNV2	O9rnv2 pseudomonas
37	29	72.5	279	5	O62416	O62416 caenorhabdi
38	29	72.5	292	10	Q9FJ16	Q9fj16 arabidopsi
39	29	72.5	296	10	Q9FRN3	Q9frn3 arabidopsi
40	29	72.5	334	10	Q9SPL2	Q9spl2 arabidopsi
41	29	72.5	389	1	Q9HR27	Q9hr27 halobacteri
42	29	72.5	405	2	Q55475	Q55475 synecocyst
43	29	72.5	406	2	O9ZJ14	O9zj14 bacillus am
44	29	72.5	424	14	Q9YZA5	Q9yza5 gallid herp
45	29	72.5	599	5	Q9VY22	Q9vy22 drosophila

ALIGNMENTS

RESULT 1

Q9JLP9 ID Q9JLP9 PRELIMINARY; PRT; 469 AA.
AC Q9JLP9;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE PRESENILIN-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SAM P8; TISSUE=HIPPOCAMPUS;
RA Kumar V.B., Vyas K.C., Choudhary V., Franko M., Flood J.F.,
RA Morley J.E.;
RT "Molecular Cloning and Tissue Distribution of Presenilin-1 in
RT Senescence Accelerated Mice (SAM P8) Mice.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF149111; AAF73153.1; -
DR InterPro; IPR001108; -
DR Pfam; PF01080; Presenilin; 1.
DR PRINTS; PR01072; PRESENILIN.
SQ SEQUENCE 469 AA; 52929 MW; CF92C2A6F398B1DF CRC64;

Query Match 90.0%; Score 36; DB 11; Length 469;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VNMAGECD 8
Db 297 VNMAGECD 303

RESULT 2

P71654 ID P71654 PRELIMINARY; PRT; 562 AA.
AC P71654;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)

Fri Sep 7 10:58:45 2001

01-JUN-2000 (TREMBlrel. 14, Last annotation update)
 HYPOTHETICAL 58.8 KDA PROTEIN.
 RV2797C OR WTCY16B7.46.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
 Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
 Taylor K., Whitehead S., Barrell B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 DR EMBL; Z81331; CAB03649.1; -
 DR TuberculList; RV2797C; -
 DR InterPro; IPR000379; -
 DR InterPro; IPR000734; -
 DR PROSITE; PS00120; LIPASE_SER; UNKNOWN_1.
 KW Hypochemical protein.
 SQ SEQUENCE 562 AA; 58791 MW; 5F6B7A6D14F9499F CRC64;

Query Match 85.0%; Score 34; DB 2; Length 562;
 Best Local Similarity 75.0%; Pred. No. 28;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVNMAEGD 8

Db 158 AVNMADGD 165

RESULT 3
 Q9XCD6 PRELIMINARY; PRT; 71 AA.

ID Q9XCD6
 AC Q9XCD6;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE HYPOTHETICAL 7.4 KDA PROTEIN.
 OS Mycobacterium smegmatis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1772;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=MC2-155;
 RX MEDLINE=99328972; PubMed=10400584;
 RA Fernandes N.D., Wu Q.-L., Kong D., Puyang X., Garg S., Husson R.N.;
 RA "A mycobacterial extracytoplasmic sigma factor involved in survival
 RT following heat shock and oxidative stress."
 RL J. Bacteriol. 181:4266-4274(1999).
 DR EMBL; AF144091; ABA41812.1; -
 DR HSSP; P10802; 11YU.
 DR InterPro; IPR000089; -
 DR Pfam; PF00364; biotin_lipoyl; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 71 AA; 7438 MW; 0E034F0263C7165E CRC64;

Query Match 80.0%; Score 32; DB 2; Length 71;
 Best Local Similarity 85.7%; Pred. No. 7.9;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VNVMAEGD 8

Db 53 VNVMAEGD 59
 RESULT 4
 ID Q27109 PRELIMINARY; PRT; 292 AA.
 AC Q27109;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE CYSTEINE PROTEINASE, PUTATIVE (FRAGMENT).
 GN CP4.
 OS Trichomonas vaginalis.
 OC Eukaryota; Parabasalida; Trichomonadida; Trichomonas.
 OX NCBI_TaxID=5722;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=G3;
 RX MEDLINE=95093621; PubMed=8000542;
 RA Mallinson D.J., Lockwood B.C., Coombs G.H., North M.J.;
 RT "Identification and molecular cloning of four cysteine proteinase
 RT genes from the pathogenic protozoan Trichomonas vaginalis."
 RL Microbiology 140:2725-2735(1994).
 DR EMBL; X77221; CAA54438.1; -
 DR HSSP; P07711; 1CJL.
 DR MEROPS; C01.082; -
 DR InterPro; IPR000169; -
 DR InterPro; IPR000668; -
 DR Pfam; PF00112; Peptidase_C1; 1.
 DR PRINTS; PR00705; PAPAIN.
 DR PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
 DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
 KW Hydrolase; Thiol protease.
 FT NON_TER 1
 FT CHAIN 75 292 CYSTEINE PROTEINASE, PUTATIVE.
 FT SEQUENCE 292 AA; 32280 MW; 862F69579BD63B45 CRC64;

Query Match 80.0%; Score 32; DB 5;
 Best Local Similarity 85.7%; Pred. No. 38;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VNVMAEGD 8

Db 187 VNVMAEGD 193

RESULT 5
 ID O60121 PRELIMINARY; PRT; 354 AA.
 AC O60121;
 DT 01-AUG-1998 (TREMBlrel. 07, Created)
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE HYPOTHETICAL 39.3 KDA PROTEIN C16G5.07C IN CHROMOSOME II.
 GN SPBC16G5.07C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OX NCBI_TaxID=4896;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Lyne M., Wood V., Rajandream M.A., Barrell B.G., Beck A.,
 Reinhardt R.;
 RA Submitted (May-1998) to the EMBL/GenBank/DBJ databases.
 RL "SIMILARITY: BELONGS TO THE BAND 7 / MEC-2 FAMILY.
 CC EMBL; AL023554; CAA19027.1; -
 DR InterPro; IPR001107; -
 DR InterPro; IPR001972; -
 DR Pfam; PF01145; Band_7; 1.

Query Match 80.0%; Score 32; DB 2; Length 71;
 Best Local Similarity 85.7%; Pred. No. 7.9;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

DR PROSITE; PS01270; BAND_7; FALSE_NEG.
 DR SMART; SMO0244; PHB; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 354 AA; 39274 MW; 66B09E6A12BDC030 CRC64;

Query Match 80.0%; Score 32; DB 3; Length 354;
 Best Local Similarity 62.5%; Pred. No. 47;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SYNMAEGD 8
 ::::|||||
 Db 229 AINVAEGD 236

RESULT 6

ID Q9SMC2 PRELIMINARY; PRT; 449 AA.
 AC Q9SMC2;
 DT 01-MAY-2000 (TRENBLrel. 13, Created)
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
 DT 01-JUN-2000 (TRENBLrel. 14, Last annotation update)
 DE ACETOLACTATE SYNTHASE SMALL SUBUNIT.
 OS Nicotiana plumbaginifolia (Leadwort-leaved tobacco).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
 OC Solanales; Solanaceae; Nicotiana.
 OX NCBI_TaxID=4092;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-99415136; PubMed=10487214;
 RA Hershey H.P., Schwartz L.J., Gale J.P., Abell L.M.;
 RT "Cloning and functional expression of the small subunit of
 RT acetolactate synthase from Nicotiana plumbaginifolia.";
 RL Plant Mol. Biol. 40:795-806(1999).
 DR EMBL; AJ234901; CAB56614.1; -;
 DR InterPro; IPR000531; -;
 DR InterPro; IPR002912; -;
 DR Pfam; PF01842; ACT; 2.
 DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.
 SQ SEQUENCE 449 AA; 50112 MW; DA0A826F37BCAD93 CRC64;

Query Match 80.0%; Score 32; DB 10; Length 449;
 Best Local Similarity 75.0%; Pred. No. 61;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SYNMAEGD 8
 ::::|||||
 Db 217 SMSMAEGD 224

RESULT 7

ID Q9S341 PRELIMINARY; PRT; 797 AA.
 AC Q9S341;
 DT 01-MAY-2000 (TRENBLrel. 13, Created)
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
 DT 01-JUN-2000 (TRENBLrel. 14, Last annotation update)
 DE OUTER MEMBRANE ANTIGEN.
 GN OMA.
 OS Photorhabdus luminescens (Xenorhabdus luminescens).
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Photorhabdus.
 OX NCBI_TaxID=29488;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HM;
 RA Chatonnet-Marton P.I., Givaudan A., Lanolis A., Boemare N.E.;
 RT "Photorhabdus luminescens genomic region homologous to 4.0 minute
 RT Escherichia coli region promotes pleiotropic phenotypes.";
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ236920; CAB51929.1; -;

DR InterPro; IPR000184; -;
 DR Pfam; PF01103; Bac_surface_Ag; 1.
 SQ SEQUENCE 797 AA; 88713 MW; 39E72E7ABD9C14F3 CRC64;

Query Match 80.0%; Score 32; DB 2; Length 797;
 Best Local Similarity 62.5%; Pred. No. 1.2e+02;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SYNMAEGD 8
 ::::|||||
 Db 257 TINMTEGD 264

RESULT 8

ID Q9MAU0 PRELIMINARY; PRT; 1064 AA.
 AC Q9MAU0;
 DT 01-OCT-2000 (TRENBLrel. 15, Created)
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
 DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
 DE F13M7.9 PROTEIN.
 GN F13M7.9.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Vysotskaia V.S., Schwartz J.R., Toriumi M., Yu G., Kwan A., Liu, S,
 RA Li J., Kremenetskaia I., Luros J., Araujo R., Au M., Bredele V.,
 RA Buehler E., Conway A., Dewar K., Feng J., Kim C., Kurtz D., Li Y.,
 RA Palm C., Shinn P., Sun H., Davis R., Ecker J., Federspiel N.,
 RA Theologis A.;
 RT "The sequence of BAC F13M7 from Arabidopsis thaliana chromosomes 1.";
 RN Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Theologis A.;
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Theologis A.;
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Theologis;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC004809; AAF40445.1; -;
 DR InterPro; IPR001296; -;
 DR Pfam; PF00534; Glycos_transf_1; 1.
 SQ SEQUENCE 1064 AA; 119675 MW; 12D97744B1B2BCFE CRC64;

Query Match 80.0%; Score 32; DB 10; Length 1064;
 Best Local Similarity 75.0%; Pred. No. 1.6e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SYNMAEGD 8
 :||:||||
 Db 32 SYNLEGD 39

RESULT 9

ID O44360 PRELIMINARY; PRT; 55 AA.
 AC O44360;
 DT 01-JUN-1998 (TRENBLrel. 06, Created)

DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DE 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE CYTOCHROME P450 MONOOXYGENASE (DU154CC) (FRAGMENT).
GN CYP6A11.
OS Ceratitis capitata (Mediterranean fruit fly).
OC Pterygota; Neoptera; Endopterygota; Tracheata; Hexapoda; Insecta;
OC Tephritidae; Tephritidae; Ceratitis.
OX NCBI_TaxID=7213;
RN [1]
RP SEQUENCE FROM N.A.
RA Danielson P.B., Foster J.L.M., Cooper S.K., Fogleman J.C.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) -> ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AF028002; AAB94119.1; -. Electron transport; Membrane; Heme;
KW Oxidoreductase; Monooxygenase;
FT NON_TER 1
SQ SEQUENCE 55 AA; 6332 MW; D182E26FF3850513 CRC64;
Query Match 77.5%; Score 31; DB 5; Length 55;
Best Local Similarity 85.7%; Pred. No. 10;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 VNMAGEGD 8
Db I:|||||
39 VSMAGEGD 45
RESULT 10
Q9H018 PRELIMINARY; PRT; 205 AA.
AC Q9H018.
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DE 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 22.8 KDA PROTEIN (FRAGMENT).
GN DKFZP667G248.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Duesterhoeft A., Lauber J., Mewes H.W., Weil B., Wiemann S.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL512767; CAC21684.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 205 AA; 22768 MW; A31C1CDF2142A8A6 CRC64;
Query Match 77.5%; Score 31; DB 4; Length 205;
Best Local Similarity 75.0%; Pred. No. 43;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 SVNMAEGD 8
Db I:|||||
35 SSNMADGD 42
RESULT 11
Q9NVT9 PRELIMINARY; PRT; 282 AA.
ID Q9NVT9
AC Q9NVT9;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)

DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE CDNA FLJ10511 FIS, CLONE NT2RP2000656.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Sugawara M.,
RA Wagatsuma M., Hosokiri T., Kaku Y., Kodaira H., Kondo H., Takiguchi S.,
RA Takahashi M., Chiba Y., Ishida S., Murakami K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y.,
RA Ninomiya K., Iwavanagi T.;
RT "NEDO human cDNA sequencing project."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK001373; BAA91656.1; -.
DR InterPro; IPR000225; -.
DR Pfam; PF00514; Armadillo_seg.1.
SQ SEQUENCE 282 AA; 31280 MW; 9CDF2DCFE6C3A08 CRC64;
Query Match 77.5%; Score 31; DB 4; Length 282;
Best Local Similarity 75.0%; Pred. No. 61;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 SVNMAEGD 8
Db I:|||||
112 SSNMADGD 119
RESULT 12
Q9HML4 PRELIMINARY; PRT; 300 AA.
ID Q9HML4
AC Q9HML4.
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE VNG2488C.
GN VNG2488C.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leithausen B., Kellner K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Eberhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT "Genome sequence of Halobacterium species NRC-1."
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL; AE005126; AAC20557.1; -.
SQ SEQUENCE 300 AA; 31668 MW; C203A3E05D8A05A9 CRC64;
Query Match 77.5%; Score 31; DB 1; Length 300;
Best Local Similarity 75.0%; Pred. No. 66;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 SVNMAEGD 8
Db I:|||||
42 SVDVAEGD 49
RESULT 13
Q9F6W5

ID Q9F6W5 PRELIMINARY; PRT; 377 AA.
 AC Q9F6W5;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE HYPOTHETICAL 41.6 KDA PROTEIN (FRAGMENT).
 OS Chloroflexus aurantiacus.
 OC Bacteria; Green non-sulfur bacteria; Chloroflexaceae group;
 CC Chloroflexaceae; Chloroflexus.
 OX NCBI_TaxID=1108;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20433268; PubMed=10976061;
 RA Xiong J., Fischer W.M., Inoue K., Nakahara M., Bauer C.E.;
 RT "Molecular evidence for the early evolution of photosynthesis.";
 RL Science 289:1724-1730(2000).
 DR EMBL; AF288461; AAG15220.1; -.
 KW Hypothetical protein.
 FT NON_TER 377 377
 SQ SEQUENCE 377 AA; 41583 MW; 6F6849081F0E98F8 CRC64;

Query Match 77.5%; Score 31; DB 2; Length 377;
 Best Local Similarity 85.7%; Pred. No. 85;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 VNMAGEG 8
 |||||
 Db 211 VNMAGD 217

RESULT 14
 Q9LLP3 PRELIMINARY; PRT; 403 AA.
 ID Q9LLP3
 AC Q9LLP3;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE HYPOTHETICAL 45.8 KDA PROTEIN.
 GN DUPR11.10.
 OS Oryza sativa (Rice).
 CC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 CC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
 CC Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20181732; PubMed=10715324;
 RA Tarchini R., Biddle P., Wineland R., Tingey S., Rafalski A.;
 RT "The complete sequence of 340 kb of DNA around the rice Adh1-adh2
 region reveals interrupted collinearity with maize chromosome 4.";
 RL Plant Cell 12:381-391(2000).
 DR EMBL; AF172282; AAF34420.1; -.
 DR InterPro; IPR001810; -.
 DR Pfam; PF00646; F-box; 1.
 DR PROSITE; PS0181; FBOX; 1.
 DR SMART; SM00256; FBOX; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 403 AA; 45782 MW; 4E943A41BE340DB4 CRC64;

Query Match 77.5%; Score 31; DB 10; Length 403;
 Best Local Similarity 62.5%; Pred. No. 91;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SYNMAEGD 8
 |:::|
 Db 99 SIDLAEGD 106

RESULT 15
 Q9P0W8 PRELIMINARY; PRT; 559 AA.
 ID Q9P0W8

AC Q9P0W8;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE HYPOTHETICAL 63.4 KDA PROTEIN (FRAGMENT).
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=TESTIS;
 RA Wang L., Miao S., Liu H., Zhang X., Ying H., Gou D.;
 RT "A full-length cDNA from Human testis cDNA library.";
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF144487; AAF66077.1; -.
 KW Hypothetical protein.
 FT NON_TER 1
 SQ SEQUENCE 559 AA; 63380 MW; 51673D79AF6EABF6 CRC64;

Query Match 77.5%; Score 31; DB 4; Length 559;
 Best Local Similarity 75.0%; Pred. No. 1.3e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SYNMAEGD 8
 |||||
 Db 497 SYNVIEGD 504

Search completed: September 6, 2001, 16:49:48
 Job time: 732 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:37:31 ; Search time 225.25 Seconds
(without alignments)
131.341 Million cell updates/sec

Title: US-09-603-713-2

Perfect score: 2587

Sequence: 1 AGVLPARGTGHGRLPLRSG.....CLRLCROHDFADDSLLK 488

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- A_Geneseq_0601.*
- 1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.*
 - 2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.*
 - 3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.*
 - 4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.*
 - 5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.*
 - 6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.*
 - 7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.*
 - 8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.*
 - 9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.*
 - 10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.*
 - 11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.*
 - 12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.*
 - 13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.*
 - 14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.*
 - 15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.*
 - 16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.*
 - 17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.*
 - 18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.*
 - 19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.*
 - 20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.*
 - 21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.*
 - 22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2587	100.0	488	22 AAB66572	Human memapsin 2.
2	2582	100.0	488	22 AAB61334	Memapsin 2 protein
3	2587	100.0	501	21 AAY94767	Human beta-secretase
4	2587	100.0	501	21 AAB07896	Amino acid sequence
5	2587	100.0	503	22 AAB66573	Human pro-memapsin
6	2587	100.0	503	22 AAB61335	T7 promoter and ve
7	2582	99.8	501	21 AAY88425	Human aspartyl pro
8	2581	99.8	501	19 AAW59807	Amino acid sequence
9	2519	97.4	501	21 AAY94769	Rat beta-secretase
10	2517	97.3	501	21 AAY94768	Murine beta-secretase
11	2517	97.3	501	21 AAY88427	Murine aspartyl pr

12	2432.5	94.0	476	21	AA88426	Human aspartyl pro
13	2419	93.5	456	21	AA807897	Active enzyme port
14	2328	90.0	453	21	AA88438	Modified human asp
15	2328	90.0	459	21	AA88439	Modified human asp
16	2324	89.8	460	21	AA807898	Amino acid sequenc
17	2320	89.7	790	19	AAW59808	Partial amino acid
18	2280	88.1	433	21	AA88433	Human-pro-Asp-2(a)
19	2280	88.1	446	21	AA88431	T7-caspase-human-p
20	2280	88.1	459	21	AA88432	T7-caspase-human-p
21	2156	83.3	415	21	AA807899	Amino acid sequenc
22	2096	81.0	425	21	AA88437	Human Asp2 amino a
23	1673	64.7	351	20	AA835918	Extended human sec
24	1175	45.4	518	19	AAW61362	Aspartic proteinase
25	1175	45.4	518	20	AA841714	Human PROB52 prote
26	1175	45.4	518	20	AA822329	Human CSP56, aspar
27	1175	45.4	518	20	AA813799	Human aspartyl pro
28	1175	45.4	518	21	AA844270	Human PROB52 (UNQ4
29	1175	45.4	518	21	AA88424	Human aspartyl pro
30	1077.5	41.7	423	22	AA88479	Human membrane of
31	296.5	11.5	412	16	AA874207	Human death associ
32	296.5	11.5	412	19	AAW71369	Death associated p
33	296.5	11.5	412	20	AAW06478	Human tumour-associ
34	296.5	11.5	412	21	AA893685	Amino acid sequenc
35	287.5	11.1	381	13	AA820730	Prochymosin (prore
36	282.5	10.9	326	22	AA866589	Human pepsin. Hom
37	282.5	10.9	326	22	AA861351	Pepsin protein. H
38	278	10.7	391	20	AA832056	Sequence encoded b
39	277.5	10.7	365	11	AA805080	Sequence of calf p
40	277.5	10.7	375	5	AA840078	Sequence encoded b
41	277.5	10.7	375	5	AA840078	Pre-prorennin-A pr
42	277.5	10.7	380	3	AA820038	Sequence of a poly
43	277.5	10.7	381	5	AA840359	Sequence of rennin
44	276.5	10.7	381	5	AA840218	prochymosin. AAP
45	273.5	10.6	365	10	AA894144	

ALIGNMENTS

RESULT 1

AAB66572

ID AAB66572 standard; Protein; 488 AA.

AC AAB66572;

DT 12-APR-2001 (first entry)

DE Human memapsin 2.

Human; memapsin 2; nototropic; neuroprotective; amyloid precursor protein; APP; memapsin 2 inhibitor; Alzheimer's disease.

OS Homo sapiens.

PN WO200100665-A2.

XX 04-JAN-2001.

XX 27-JUN-2000; 2000WO-US17742.

XX 28-JUN-1999; 99US-0141363.

XX 30-NOV-1999; 99US-0168060.

XX 25-JAN-2000; 2000US-017836.

XX 27-JAN-2000; 2000US-0178368.

XX 08-JUN-2000; 2000US-0210292.

XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.

XX (UNII) UNIV ILLINOIS FOUND.

XX Tang JJN, Hong L, Ghosh AK;

XX WPI: 2001-137933/14.

XX N-PSDB; AAF1848.

XX PT - Novel memapsin 2 inhibitors which bind to active site of memapsin 2
XX having 2 catalytic aspartic residues and substrate binding cleft, used
XX to treat Alzheimer's disease by blocking amyloid precursor protein
XX cleavage
XX
XX Example 1; Page 72-74; 86pp; English.
XX
XX The present sequence is given in a specification relating to an inhibitor
XX of catalytically active memapsin 2. The inhibitor binds to the memapsin 2
XX active site, which is defined by the presence of two catalytic aspartic
XX residues and a substrate binding cleft. The inhibitor is useful for
XX the treatment and diagnosis of Alzheimer's disease. It is useful in
XX screens for individuals with a genetic predisposition to Alzheimer's
XX disease. The inhibitor is useful as a reagent for specifically binding to
XX memapsin 2 or memapsin 2 analogues and for aiding in memapsin 2
XX isolation, purification and characterisation.
XX
XX Sequence 488 AA;
XX
XX Query Match 100.0%; Score 2587; DB 22; Length 488;
XX Best Local Similarity 100.0%; Pred. No. 4.4e-257;
XX Matches 488; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 AGVLPAGHTQHGIRLPLRSGLGAPLGLRLPRETDEEPEEGRGSGFVEMVDNLRGKSGQ 60
XX Db 1 agvlpaghtqhgirplrslggagplgrlpretdeepeegrrgsfvmvndnlrgksqg 60
XX
XX QY 61 GYVEMTVGSPPTNLILVDGTSSNFVAGAAPHPFLHRYQRLSTYRDLRKGVVYPT 120
XX Db 61 gyyvemtvgspptnlilvdtgssnfavgaaphflhryyqrlsyttyrldrkgyvpyt 120
XX
XX QY 121 QGKWEGLGTDLVSIHPGNVTVRANIAAITESDKFFINGSNWEGILGLAYAEIARPD 180
XX Db 121 qgkwegeigtldvsihpgnvtvraniaaitesdkffingsnwegilglayaeiarpdds 180
XX
XX QY 181 LEPPFDSLKVQTHVPNLFSQLCGAGFPLNQSEVLASVGGSMIIGGDHSLYTGSLWYTP 240
XX Db 181 leppfdsldkvthvpnlfsqlcgagfplngsevlavsggsmliiggdhslytgslywtp 240
XX
XX QY 241 IRREWYVEIIVRVEINGODLKMCKEYNDKSIIVDSGTTNLRPKKVFEEAAVKSIAAS 300
XX Db 241 irrewyveiiivrveingodlkmckeynydksivdsgettlnlrpkkvfeaaavksikaas 300
XX
XX QY 301 STEKFPDGFWLGEOLVCWQAGTTPWNIFPVISLYLMGEVNTNOSFRITILPQOYLRPVEDV 360
XX Db 301 stekfpdgfwlgeolvcwqagttwnifpvvislylmgevtngsfrtilpqylrpvedv 360
XX
XX QY 361 ATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVDRAKRIGFAVSACHVHDEFRTAAVEG 420
XX Db 361 atsqqddcykfaisqsstgtvmgavimegyvvdarakrigfavsaachvhdefrtaaveg 420
XX
XX QY 421 PFVTLDMEDCGYNIPQTDSTLMTIAYVMAAICALFMLEPLCMVCWRCRLCRLRQHHDF 480
XX Db 421 pfvtlmedcgyinipqtdstlmtiayvmaaaicalfmlplcmvcwrcrlcrlrqhddf 480
XX
XX QY 481 ADDISLLK 488
XX Db 481 addisllk 488
XX
XX RESULT 2
XX AAB61334
XX ID AAB61334 standard; protein; 488 AA.
XX XX
XX AC AAB61334;
XX XX
XX DT 02-APR-2001 (first entry)
XX XX
XX DE Memapsin 2 protein.
XX XX
XX KW Memapsin 2; catalyst; Alzheimer's.

XX OS Homo sapiens.
XX PN WO200100663-A2.
XX PD 04-JAN-2001.
XX PF 27-JUN-2000; 2000WO-US17661.
XX PR 28-JUN-1999; 99US-0141363.
XX PR 30-NOV-1999; 99US-0168060.
XX PR 25-JAN-2000; 2000US-017836.
XX PR 27-JAN-2000; 2000US-0178368.
XX PR 08-JUN-2000; 2000US-0210292.
XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.
XX PA Tang JUN, Lin X, Koelsch G;
XX PI WPI; 2001-102885/11.
XX DR Purified recombinant catalytically active memapsin 2, used to screen
XX inhibitors of it, which are used to treat and prevent Alzheimer's
XX disease -
XX Claim 2; Page 73-75; 86pp; English.
XX The present invention relates to a purified recombinant
XX catalytically active memapsin 2. The invention may be used for
XX isolating inhibitors which are used to treat or prevent
XX Alzheimer's disease. The invention may also be used to screen
XX for individuals more genetically prone to develop Alzheimer's
XX disease.
XX Sequence 488 AA;
XX
XX Query Match 100.0%; Score 2587; DB 22; Length 488;
XX Best Local Similarity 100.0%; Pred. No. 4.4e-257;
XX Matches 488; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 AGVLPAGHTQHGIRLPLRSGLGAPLGLRLPRETDEEPEEGRGSGFVEMVDNLRGKSGQ 60
XX Db 1 agvlpaghtqhgirplrslggagplgrlpretdeepeegrrgsfvmvndnlrgksqg 60
XX
XX QY 61 GYVEMTVGSPPTNLILVDGTSSNFVAGAAPHPFLHRYQRLSTYRDLRKGVVYPT 120
XX Db 61 gyyvemtvgspptnlilvdtgssnfavgaaphflhryyqrlsyttyrldrkgyvpyt 120
XX
XX QY 121 QGKWEGLGTDLVSIHPGNVTVRANIAAITESDKFFINGSNWEGILGLAYAEIARPD 180
XX Db 121 qgkwegeigtldvsihpgnvtvraniaaitesdkffingsnwegilglayaeiarpdds 180
XX
XX QY 181 LEPPFDSLKVQTHVPNLFSQLCGAGFPLNQSEVLASVGGSMIIGGDHSLYTGSLWYTP 240
XX Db 181 leppfdsldkvthvpnlfsqlcgagfplngsevlavsggsmliiggdhslytgslywtp 240
XX
XX QY 241 IRREWYVEIIVRVEINGODLKMCKEYNDKSIIVDSGTTNLRPKKVFEEAAVKSIAAS 300
XX Db 241 irrewyveiiivrveingodlkmckeynydksivdsgettlnlrpkkvfeaaavksikaas 300
XX
XX QY 301 STEKFPDGFWLGEOLVCWQAGTTPWNIFPVISLYLMGEVNTNOSFRITILPQOYLRPVEDV 360
XX Db 301 stekfpdgfwlgeolvcwqagttwnifpvvislylmgevtngsfrtilpqylrpvedv 360
XX
XX QY 361 ATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVDRAKRIGFAVSACHVHDEFRTAAVEG 420
XX Db 361 atsqqddcykfaisqsstgtvmgavimegyvvdarakrigfavsaachvhdefrtaaveg 420
XX
XX QY 421 PFVTLDMEDCGYNIPQTDSTLMTIAYVMAAICALFMLEPLCMVCWRCRLCRLRQHHDF 480
XX Db 421 pfvtlmedcgyinipqtdstlmtiayvmaaaicalfmlplcmvcwrcrlcrlrqhddf 480

QY 481 ADDISLLK 488
 Db 481 addisllk 488

RESULT 3
 AAY94767
 ID AAY94767 standard; Protein: 501 AA.
 XX
 AC AAY94767;
 XX
 DT 12-FEB-2001 (first entry)
 DE Human beta-secretase amino acid sequence.
 XX
 KW Beta-secretase; enzyme; amyloid plaque; Alzheimer's disease; human;
 KW Down's syndrome; amyloid angiopathy; gene therapy; neuroprotective.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..45
 FT /label= putative signal peptide
 FT Protein 46..501
 FT /label= Beta-secretase
 XX
 PN WO200058479-A1.
 XX
 PD 05-OCT-2000.
 XX
 PF 23-MAR-2000; 2000WO-US07755.
 XX
 PR 26-MAR-1999; 99US-0277229.
 XX
 PA (AMGE-) AMGEN INC.
 XX
 PI Citron M, Vassar RJ, Bennett BD;
 XX
 DR WPI: 2000-594643/56.
 DR N-PSDB; AAA28278.
 XX
 PT Isolated beta-secretase nucleic acids and encoded polypeptides, useful
 PT for diagnosis and gene therapy of Alzheimer's disease -
 XX
 PS Claim 1; Fig 4; 145pp; English.
 XX

CC This invention relates to 3 nucleotide sequences encoding beta-secretase
 CC proteins. Beta-secretase is an enzyme involved in the production of one
 CC of the components of amyloid plaques involved in Alzheimer's disease. The
 CC invention includes an expression vector comprising the nucleotide
 CC sequence, a host cell comprising the expression vector, and a process for
 CC producing the protein through culturing the transformed cells. Also
 CC included in the invention are a polypeptide derivative of the
 CC beta-secretase protein, a fusion protein comprising beta-secretase fused
 CC to a heterologous amino acid sequence, and a method for modulating the
 CC levels of beta-secretase polypeptide in a mammal comprising administering
 CC the polynucleotide sequence. Beta-secretase exhibits neuroprotective and
 CC neurotropic activity. The beta-secretase nucleotide sequence may be used to
 CC map locations of the beta-secretase gene and related genes on chromosomes
 CC and as hybridization probes in diagnostic assays to test for the presence
 CC of beta-secretase DNA or RNA, such as in Alzheimer's disease, Down's
 CC syndrome, and amyloid angiopathy. The nucleotide sequence may also be
 CC used as anti-sense inhibitors of beta-secretase expression, in gene
 CC therapy of Alzheimer's disease, and for the identification of compounds
 CC that modulate beta-secretase activity. Antibodies to the beta-secretase
 CC protein may be used for in vitro and in vivo diagnostic purposes to
 CC detect the presence of beta-secretase polypeptide in a body fluid or cell
 CC sample. The present sequence represents the human beta-secretase protein.
 XX

SQ Sequence 501 AA;

Query Match 100.0%; Score 2587; DB 21; Length 501;

Best Local Similarity 100.0%; Pred. No. 4.6e-257;
 Matches 488; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGVLPAGHTQHIGIRLPRLRSGLGAPLGLRLPRETDEPEEPGRGSRFVEMVNLKRSQ 60
 |||||
 Db 14 agvlpahgtqhigiriprlrsglgagplglrlpretdeepeepgrgrsfvemdnlrgksq 73
 |||||
 QY 61 GYVEMTVGSPPTNLNLYDVTGSSNFACAPHPFLHRYRQRLSTYRDLRKGVVYPT 120
 |||||
 Db 74 gyyvemtvgspptnlnlydvtgssnfavgaaphpflhryyqrqlstyrdlrgkvvypyt 133
 |||||
 QY 121 QGKWEGLTGLVSIPIHGPNTVRANIAAITESDKFFINGSNWEGILGLAYAEIARPD 180
 |||||
 Db 134 qgkwegelgtldlvsiphgpnvtvranaiaatesdkffingsnwegilglayaeiarpd 193
 |||||
 QY 181 LEPFFDSLKQTHVPLNLSLQCGAGFPLNQSEVLASVSGSMIIGIDHSITGSLWYTP 240
 |||||
 Db 194 lepffdsllvkqthvpnlfsllqcgagfplnqsevlavsgsgmligldhsltytgslywtp 253
 |||||
 QY 241 IRREWYEVIIIVRVEINGQDLKMDCKEYNDKSIYDVGTTNLRPKKVFEEAAVKSIKAAS 300
 |||||
 Db 254 irrewyevilvrveingqdlkmdckeyndksiydvgttnlrlpkkvfeavksikaas 313
 |||||
 QY 301 STEKPPDGFMLGEQLVCMQAGTTPWNIFPVISLYLMGEVTNOSFRITILPQOYLRPVEDV 360
 |||||
 Db 314 stekppdggfwlgeqlvcmqagttwnifpvlslylmgevtngsfrtililpqylrpvedv 373
 |||||
 QY 361 ATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSACHVHDEFRTAAVEG 420
 |||||
 Db 374 atsqddcykfaisqsgstgtvmgavimegyvfvdrarkrigrfavsachvhdefrtaaveg 433
 |||||
 QY 421 PFVTLDMEDCGYNIPTDESTLMTIAYVMAATCALPMLPCLMVCQWRCGLRQOHD 480
 |||||
 Db 434 pvtldmedcgynipqtdestlmtiayvmaatcalmfpclmvcqwrclrlrqhndf 493
 |||||
 QY 481 ADDISLLK 488
 |||||
 Db 494 addisllk 501

RESULT 4
 AAB07896
 ID AAB07896 standard; Protein: 501 AA.
 XX
 AC AAB07896;
 XX
 DT 14-NOV-2000 (first entry)
 DE Amino acid sequence of a human beta-secretase enzyme.
 XX
 KW Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;
 KW amyloid plaque component; Alzheimer's disease; amyloidogenic disease;
 KW inhibitor.
 XX
 OS Homo sapiens.
 XX
 PN WO200047618-A2.
 XX
 PD 17-AUG-2000.
 XX
 PF 10-FEB-2000; 2000WO-US03819.
 XX
 PR 10-FEB-1999; 99US-0119571.
 PR 15-JUN-1999; 99US-0139172.
 XX
 PA (ELAN-) ELAN PHARM INC.
 XX
 PI Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;
 PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;
 XX
 DR WPI: 2000-533011/48.
 DR N-PSDB; AAA59550, AAA59551.
 XX

RESULT	5	
AA866573		
ID	AA866573 standard; Protein; 503 AA.	
XX		
AC		
AA866573;		
XX		
XX		
DT		
12-APR-2001	(first entry)	
XX		
DE		
Human pro-memapsin 2.		
XX		
Human; memapsin 2; neurotropic; amyloid precursor protein;		
KW		
APP; memapsin 2 inhibitor; Alzheimer's disease; ss.		

Query Match	100.0%;	Score 2587;	DB 22;	Length 503;
Best Local Similarity	100.0%;	Pred. No. 4.7e-257;		
Matches 488;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	AGVLPAHGTOHGIRLPURSGIGGAPLGLRLPREWDEPEEPGRGSGFVEMVDNLRKSGQ	60	
Db				
Db	16	agvlpahgtghrlrlpsrglgsppglglrlpredeepeepgrrgsfvemvohnlrksgq	75	
QY	61	GYVEMTVGSPQPTNLIVDTGSSNFAVGAAPHFLHRYYQRLSSYTRDLRKGIVVPYT	120	
Db				
Db	76	gyvemtvgspqptnlilvdtgssnfavgaaphflhryyqrlssytrdlrkgyvpyt	135	
QY	121	QKWEGLGTDLYSIPHGPNVTVRNTAAITESDKFFINGSNWEGILGLAYEARTPDDS	180	
Db				
Db	136	qkwegeigtclvshpghnvtvrantaatesdkffingsnwegilglayaeartpdds	195	
QY	181	LEPFFSLVKQTHVPNLFLSLQLCGAGPPLNQSEVLASVGGSMITGGIDHSYTGSLWTP	240	
Db				
Db	196	lepffslvkqthvnpnlflslqlcgagpplnqsevlasvgsmltggidhslytgslywtp	255	
QY	241	IRREWYEVYIIRVEINGDLKMDCKEYNDKSTVDSGTTNLRLPKPKVFEAAVKSIAAS	300	
Db				
Db	256	irrewyevyilvrveingdqlkmdckeyndkslvdsgttnlrtpkpvfeaaavksiaas	315	
QY	301	STKEFPDGFWLGEQLVCWQAGTTPNNIFPVISLYLMEVNTQSPRIILPQOYLRPVEDV	360	
Db				
Db	316	stkefpdgfwlgeqlvcwqagttpwnlfpvisllylmevntqsfriilpqoylrpvedv	375	
QY	361	ATSQDDCYKFAISQSSTGTVMGAVIMEGFVYVDFRARKRIGFAVSACHVHDEFRTAAVEG	420	
Db				
Db	376	atsqddcvkfaissgstatvmgavimegfvyvdfdrakrigrfavsachvhdefrtaveg	435	

```
OY 421 PFVTLDMEDCGYNIPQTDSTLMTIAYVMAAICALFMLPICLMLVCQWRCLRLRQHQHDDF 480
|||||
Db 436 pfvtldmedcgnipqtdstlmtiayvmaaaicalfmlpiclmlvcqwrclrlrqghddf 495
OY 481 ADDISLLK 488
Db 496 addisllk 503

RESULT 6
AAB61335
ID AAB61335 standard; protein: 503 AA.
XX
AC AAB61335;
XX
DT 02-APR-2001 (first entry)
XX
DE T7 promoter and vector sequence.
XX
KW Memapsin 2; catalyst; Alzheimer's.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200100663-A2.
XX
PD 04-JAN-2001.
XX
PF 27-JUN-2000; 2000WO-US17661.
XX
PR 28-JUN-1999; 99US-0141363.
PR 30-NOV-1999; 99US-0168060.
PR 25-JAN-2000; 2000US-0177836.
PR 27-JAN-2000; 2000US-0178368.
PR 08-JUN-2000; 2000US-0210292.
XX
PA (OKLA-) OKLAHOMA MEDICAL RES FOUND.
XX
PI Tang JJN, Lin X, Koelsch G;
DR WPI; 2001-102885/11.
XX
PT Purified recombinant catalytically active memapsin 2, used to screen
PT inhibitors of it, which are used to treat and prevent Alzheimer's
PT disease -
XX
PS Disclosure; Fig 1; 86pp; English.
XX
CC The present invention relates to a purified recombinant
CC catalytically active memapsin 2. The invention may be used for
CC isolating inhibitors which are used to treat or prevent
CC Alzheimer's disease. The invention may also be used to screen
CC for individuals more genetically prone to develop Alzheimer's
CC disease.
XX
SQ Sequence 503 AA:

Query Match 100.0%; Score 2587; DB 22; Length 503;
Best Local Similarity 100.0%; Pred. No. 4.7e-257;
Matches 488; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGVLPAHGTGIRLPLRSGLGGAPLGLRLPRETDEEPEEPRGRGSFVEMVDNLRKSGQ 60
|||||
Db 16 agvlpahgtgqirplrslggagplrlpretdeeepegrgrgsfvmvndlrksgq 75
OY 61 GYVEMTVGSPQPTNLIVDTGSSNFAGAAPFLHRYQRLSSYRDLRKGVVYPYT 120
|||||
Db 76 gyyvemtvgspqptnlilvdtgssnfavgaaphflhryyqrlssyrdlrkvvyvpyt 135
OY 121 QGKWEGLGTLDSIPGPNVTPYRNTAAITESDKFPINGSNWEGILGLAYAEIARDDDS 180
|||||
```

```
Db 136 qgkwegelgtclsvsihpnpnvtvranaiaatesdkffingsnwegilglayaeiarpdds 195
OY 181 LEFFFDLVKQTHVPNLFSLQACAGEPPLMQSEVLASVGGSMITGGIDHSLYTGSLWYTP 240
|||||
Db 196 lefffdslvkqthvpnlfsllqcgagfplnqsevlavsggsmlggidhsltytgslytp 255
OY 241 IRREWYVEVLIIVRVEINGQDLKWDCKEYNDKSIYDSTTNLRLPKVFEAAVKSIIKAAS 300
|||||
Db 256 irrewyvevliivrveingqdlkmdckeyndksivdsgttnlrlpkkvfeavksikaas 315
OY 301 STEKFPDGLWGLQVLCWQAGTTPWNIFFVTSILYLMGEVTNOSFRITILPOOYLRPVEDV 360
|||||
Db 316 stekfpdgfwlgeqlvcwqagttpnifvtsilylmgevtngsfrilitlppqylrpvedv 375
OY 361 ATSODDCYKFAISOSSTGTVMGAVIMEGFYVVDRAKRIGFAVSACHVDEEPTAAVEG 420
|||||
Db 376 atsoddcykfaissstgtvmgavimegfyvvdrrakrigfavsvachvdeftaaaveg 435
OY 421 PFVTLDMEDCGYNIPQTDSTLMTIAYVMAAICALFMLPICLMLVCQWRCLRLRQHQHDDF 480
|||||
Db 436 pfvtldmedcgnipqtdstlmtiayvmaaaicalfmlpiclmlvcqwrclrlrqghddf 495
OY 481 ADDISLLK 488
Db 496 addisllk 503

RESULT 7
AAY88425
ID AAY88425 standard; Protein: 501 AA.
XX
AC AAY88425;
XX
DT 03-AUG-2000 (first entry)
XX
DE Human aspartyl protease 2 (a) (Asp2) amino acid sequence.
XX
KW Aspartyl protease; aspartase; amyloid precursor protein; APP; Asp 2;
KW Alzheimer's disease; beta secretase site.
XX
OS Homo sapiens.
XX
PN WO200017369-A2.
XX
PD 30-MAR-2000.
XX
PF 23-SEP-1995; 99WO-US20881.
XX
PR 24-SEP-1995; 98US-0101594.
XX
PA (PHAA ) PHARMACIA & UPJOHN CO.
XX
PI Gurney ME, Bienkowski MJ, Heinrikson RL, Parodi LA, Yan R;
XX
DR WPI; 2000-303209/26.
DR N-PSDB; AAA15662.
XX
PT New enzyme designated human aspartase useful in research into
PT Alzheimer's Disease is capable of cleaving amyloid protein precursor at
PT the beta secretase site to produce amyloid beta peptide -
XX
PS Claim 48; Fig 2; 183pp; English.
XX
CC This sequence represents the human aspartyl protease 2 (Asp2) amino acid
CC sequence. The invention relates to a protease (e.g. Asp2) capable of
CC cleaving the beta secretase site of amyloid precursor protein (APP). The
CC protease contains a sequence encoding the amino acid sequence DTG and a
CC sequence encoding DSG or DTG separated by 100-300 amino acids. When
CC mutated the APP gene causes an autosomal dominant form of Alzheimer's
CC disease. APP localises to the cell surface membrane and have a single
CC C-terminal transmembrane domain. Proteolytic processing of APP produces
CC the amyloid beta protein, which is possibly very important in Alzheimer's
CC disease. The invention includes a nucleotide sequence encoding the
```

CC protease, a vector containing the nucleotide sequence, and a cell line
 CC comprising the vector. Methods for screening for inhibitors of beta
 CC secretase activity are also given in the invention. The human aspartase
 CC protein and nucleotide sequences and the methods for identifying
 CC inhibitors of the protease, are useful in the treatment of and research
 CC in to Alzheimer's disease.
 XX
 SQ Sequence 501 AA;

Query Match 99.8%; Score 2582; DB 21; Length 501;
 Best Local Similarity 99.8%; Pred. No. 1.5e-256;
 Matches 487; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGVLPAGHTQHGIRLPLRSGLGGAPLGLRLPRETDEEPEEGRGGSFVEMVDNLRKSGQ 60
 Db 14 agvlpahgtqhgirplrsglggaplgirlpretdeeepeegrsgsfvemdnlrksqg 73
 QY 61 GYVEMTVGSPQTNLILVDTGSSNFAGVGAAPHFLHRYQRLSSYRDLRKGVVYPYT 120
 Db 74 gyyvemtvgspqtnlilvdtgssnfavgaapflhryyqrlsstyrdlrkgvvypt 133
 QY 121 QGKWEGLGTDLVSIIPHGPNTVVRANIAAITESDKFFINGSNWEGILGLAYAEIARPD DS 180
 Db 134 gqkwegelgtldvsiiphgpnvtvrانياaitesdkffingsnwegilglayaelarpdds 193
 QY 181 LEPPFDSLVKQTHVPNLFSLQCGAGFPLNQSEVLASVGGSMIIGGDHSLYTGSLWYTP 240
 Db 194 lepfdslvkqthvnpnlfsllcagfplnqsevlavsggsmliggdhsltygslwytp 253
 QY 241 IRREWYEVIIIVRVEINGQDLKMDCKEYNDKSTIVDSGTTNLRPKKVFEEAAVKSIIKAAS 300
 Db 254 irrewyeviiivrveingqdlkmdckeynydkstivdsgettlnlrpkkvfeaaavksikaas 313
 QY 301 STEKFPDGFNLGEOQLVCWQAGTTPWNIFPVISLYLMGEVTVNQSFRTILPQQYLRPVEDV 360
 Db 314 stekfpdgfnlgedqlvcwqagttwnifpvlslylmgevtvnqsfritilpqqylrpvedv 373
 QY 361 ATSDDCYKFAISQSSTGTVMGAVIMEGFYVDFDRKRIGFAVSACHVHDEFRTAAVEG 420
 Db 374 atsqdcykfaigsstgtvmgavimegfyyvdfdrkrigfavsachvhdefrtaaveg 433
 QY 421 PFVTLDMEDCGYNIPQDDESTLMTIAYVMAAICALFMLPLCMVCQWRCRLCRLRQHQHDF 480
 Db 434 pfvtlmedcgynipqdestlmtiayvmaaaicalfmlplcmvcwrcrlcrlrqghddf 493
 QY 481 ADDISLLK 488
 Db 494 addisllk 501

RESULT 8
 AAW59807
 ID AAW59807 standard; Protein; 501 AA.
 XX
 AC AAW59807;

DT 26-OCT-1998 (first entry)

DE Amino acid sequence of human ASP2 (aspartic protease 2).

KW Human; ASP2; aspartic protease 2; agonist; antagonist; immunospecific;
 KW antibody; inhibition; Alzheimer's disease; cancer; proteinase;
 KW prohormone processing.

XX Homo sapiens.

XX EP855444-A2.

XX 29-JUL-1998.

XX 27-JAN-1998; 98EP-0300573.

XX

PR 28-JAN-1997; 97GB-0001684.
 XX (SMIK) SMITHKLINE BEECHAM CORP.
 PA (SMIK) SMITHKLINE BEECHAM PLC.
 XX Chapman CG, Murphy K, Powell DJ, Smith TS;
 PI WPI: 1998-389809/34.
 XX N-PSDB; AAV41696.
 DR
 DR New nucleic acid encoding human aspartic protease 2 - used to treat,
 PT prevent and diagnose e.g. Alzheimer's disease, cancer and prohormone
 PT processing
 XX
 PS Claim 1; Page 7; 26pp; English.
 XX
 CC This is the amino acid sequence of the human ASP2 (aspartic protease
 CC family), used in the method of the invention. Agonists and
 CC antagonists for ASP2 immunospecific antibodies are used to treat
 CC conditions requiring increased or decreased activity or expression of
 CC ASP2 respectively. ASP2 is used to treat and diagnose e.g.
 CC Alzheimer's disease, cancer and prohormone processing and ASP2 or a
 CC fragment can be used to induce an immune response against the above
 CC conditions.
 CC
 XX Sequence 501 AA;

Query Match 99.8%; Score 2581; DB 19; Length 501;
 Best Local Similarity 99.8%; Pred. No. 1.9e-256;
 Matches 487; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGVLPAGHTQHGIRLPLRSGLGGAPLGLRLPRETDEEPEEGRGGSFVEMVDNLRKSGQ 60
 Db 14 agvlpahgtqhgirplrsglggaplgirlpretdeeepeegrsgsfvemdnlrksqg 73
 QY 61 GYVEMTVGSPQTNLILVDTGSSNFAGVGAAPHFLHRYQRLSSYRDLRKGVVYPYT 120
 Db 74 gyyvemtvgspqtnlilvdtgssnfavgaapflhryyqrlsstyrdlrkgvvypt 133
 QY 121 QGKWEGLGTDLVSIIPHGPNTVVRANIAAITESDKFFINGSNWEGILGLAYAEIARPD DS 180
 Db 134 gqkwegelgtldvsiiphgpnvtvrانياaitesdkffingsnwegilglayaelarpdds 193
 QY 181 LEPPFDSLVKQTHVPNLFSLQCGAGFPLNQSEVLASVGGSMIIGGDHSLYTGSLWYTP 240
 Db 194 lepfdslvkqthvnpnlfsllcagfplnqsevlavsggsmliggdhsltygslwytp 253
 QY 241 IRREWYEVIIIVRVEINGQDLKMDCKEYNDKSTIVDSGTTNLRPKKVFEEAAVKSIIKAAS 300
 Db 254 irrewyeviiivrveingqdlkmdckeynydkstivdsgettlnlrpkkvfeaaavksikaas 313
 QY 301 STEKFPDGFNLGEOQLVCWQAGTTPWNIFPVISLYLMGEVTVNQSFRTILPQQYLRPVEDV 360
 Db 314 stekfpdgfnlgedqlvcwqagttwnifpvlslylmgevtvnqsfritilpqqylrpvedv 373
 QY 361 ATSDDCYKFAISQSSTGTVMGAVIMEGFYVDFDRKRIGFAVSACHVHDEFRTAAVEG 420
 Db 374 atsqdcykfaigsstgtvmgavimegfyyvdfdrkrigfavsachvhdefrtaaveg 433
 QY 421 PFVTLDMEDCGYNIPQDDESTLMTIAYVMAAICALFMLPLCMVCQWRCRLCRLRQHQHDF 480
 Db 434 pfvtlmedcgynipqdestlmtiayvmaaaicalfmlplcmvcwrcrlcrlrqghddf 493
 QY 481 ADDISLLK 488
 Db 494 addisllk 501

RESULT 9
 AAY94769
 ID AAY94769 standard; Protein; 501 AA.
 XX


```

AC AAY94769;
XX
DT 12-FEB-2001 (first entry)
XX
DE Rat beta-secretase protein.
XX
KW Beta-secretase; enzyme; amyloid plaque; Alzheimer's disease; rat;
KW Down's syndrome; amyloid angiopathy; gene therapy; neuroprotective; ss.
XX
OS Rattus sp.
XX
PN WO200058479-A1.
XX
PD 05-OCT-2000.
XX
PF 23-MAR-2000; 2000WO-US07755.
XX
PR 26-MAR-1999; 99US-0277229.
XX
PA (AMGE-) AMGEN INC.
XX
PI Citron M, Vassar RJ, Bennett BD;
XX
DR WPI: 2000-594643/56.
DR N-PSDB; AAA28280.
XX
PT Isolated beta-secretase nucleic acids and encoded polypeptides, useful
PT for diagnosis and gene therapy of Alzheimer's disease.
XX
PS Claim 1; Fig 6; 145pp; English.
XX
CC This invention relates to 3 nucleotide sequences encoding beta-secretase
CC proteins. Beta-secretase is an enzyme involved in the production of one
CC of the components of amyloid plaques involved in Alzheimer's disease. The
CC invention includes an expression vector comprising the nucleotide
CC sequence, a host cell comprising the expression vector, and a process for
CC producing the protein through culturing the transformed cells. Also
CC included in the invention are a polypeptide derivative of the
CC beta-secretase protein, a fusion protein comprising beta-secretase fused
CC to a heterologous amino acid sequence, and a method for modulating the
CC levels of beta-secretase polypeptide in a mammal comprising administering
CC the polynucleotide sequence. Beta-secretase exhibits neuroprotective and
CC nootropic activity. The beta-secretase nucleotide sequence may be used to
CC map locations of the beta-secretase gene and related genes on chromosomes
CC and as hybridization probes in diagnostic assays to test for the presence
CC of beta-secretase DNA or RNA, such as in Alzheimer's disease, Down's
CC syndrome, and amyloid angiopathy. The nucleotide sequence may also be
CC used as anti-sense inhibitors of beta-secretase expression, in gene
CC therapy of Alzheimer's disease, and for the identification of compounds
CC that modulate beta-secretase activity. Antibodies to the beta-secretase
CC protein may be used for in vitro and in vivo diagnostic purposes to
CC detect the presence of beta-secretase polypeptide in a body fluid or cell
CC sample. The present sequence represents the rat beta-secretase protein.
XX
SQ Sequence 501 AA;

```

```

Query Match 97.4%; Score 2519; DB 21; Length 501;
Best Local Similarity 96.9%; Pred. No. 4.6e-250;
Matches 473; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

```

```

QY 181 LEFFDLSLVKQTHVNPFLSGLQCGAGFPLNQSEVLASVGSGSMIIGDHSLYTGSWYTP 240
DB 194 LEFFDLSLVKQTHVNPFLSGLQCGAGFPLNQSEVLASVGSGSMIIGDHSLYTGSWYTP 253
QY 241 IRREWYVEVLIIVRVEINGODLKMDCKEYNYDKSIVDSGTTNLRPKKVFEEAAVKSIAAS 300
DB 254 IRREWYVEVLIIVRVEINGODLKMDCKEYNYDKSIVDSGTTNLRPKKVFEEAAVKSIAAS 313
QY 301 STEKFPDGFGLGQLVCMQAGTTPWNIFPVISLYLMGEVTNQSFRITILPQQYLRPVEDV 360
DB 314 STEKFPDGFGLGQLVCMQAGTTPWNIFPVISLYLMGEVTNQSFRITILPQQYLRPVEDV 373
QY 361 ATSDDCYKFAISQSSTGTVMGAVINMEGVYVDFDRARKRIGFAVSACHVHDEPRTAAVEG 420
DB 374 ATSDDCYKFAISQSSTGTVMGAVINMEGVYVDFDRARKRIGFAVSACHVHDEPRTAAVEG 433
QY 421 PFVTLDMEDCGYNIPOTDESTLMTIAYVMAAICALFMLPLCLMVCOMRCLRLCRLQHQHDDF 480
DB 434 PFVTLDMEDCGYNIPOTDESTLMTIAYVMAAICALFMLPLCLMVCOMRCLRLCRLQHQHDDF 493
QY 481 ADDISLLK 488
DB 494 ADDISLLK 501

```

```

RESULT 10
AAY94768
ID AAY94768 standard; Protein; 501 AA.
XX
AC AAY94768;
XX
DT 12-FEB-2001 (first entry)
XX
DE Murine beta-secretase amino acid sequence.
XX
KW Beta-secretase; enzyme; amyloid plaque; Alzheimer's disease; mouse;
KW Down's syndrome; amyloid angiopathy; gene therapy; neuroprotective.
XX
OS Mus sp.
XX
PN WO200058479-A1.
XX
PD 05-OCT-2000.
XX
PF 23-MAR-2000; 2000WO-US07755.
XX
PR 26-MAR-1999; 99US-0277229.
XX
PA (AMGE-) AMGEN INC.
XX
PI Citron M, Vassar RJ, Bennett BD;
XX
DR WPI: 2000-594643/56.
DR N-PSDB; AAA28279.
XX
PT Isolated beta-secretase nucleic acids and encoded polypeptides, useful
PT for diagnosis and gene therapy of Alzheimer's disease.
XX
PS Claim 1; Fig 5; 145pp; English.

```

```

CC This invention relates to 3 nucleotide sequences encoding beta-secretase
CC proteins. Beta-secretase is an enzyme involved in the production of one
CC of the components of amyloid plaques involved in Alzheimer's disease. The
CC invention includes an expression vector comprising the nucleotide
CC sequence, a host cell comprising the expression vector, and a process for
CC producing the protein through culturing the transformed cells. Also
CC included in the invention are a polypeptide derivative of the
CC beta-secretase protein, a fusion protein comprising beta-secretase fused
CC to a heterologous amino acid sequence, and a method for modulating the
CC levels of beta-secretase polypeptide in a mammal comprising administering
CC the polynucleotide sequence. Beta-secretase exhibits neuroprotective and
CC nootropic activity. The beta-secretase nucleotide sequence may be used to
CC map locations of the beta-secretase gene and related genes on chromosomes
CC

```

and as hybridization probes in diagnostic assays to test for the presence of beta-secretase DNA or RNA, such as in Alzheimer's disease, Down's syndrome, and amyloid angiopathy. The nucleotide sequence may also be used as anti-sense inhibitors of beta-secretase expression, in gene therapy of Alzheimer's disease, and for the identification of compounds that modulate beta-secretase activity. Antibodies to the beta-secretase protein may be used for *in vitro* and *in vivo* diagnostic purposes to detect the presence of beta-secretase polypeptide in a body fluid or cell sample. The present sequence represents the murine beta-secretase protein.

SQ Sequence 501 AA;

Query Match 97.3%; Score 2517; DB 21;
Best Local Similarity 96.9%; Pred. No. 7.3e-250;
Matches 473; Conservative 6; Mismatches 0;

[illegible]

RESULT 11

AA88427

ID AAY88427 standard; protein; 501 AA.

XX

AC AAY88427;

XX
DT 03-AUG-2000 (first entry)XX
XX

DE
XX
Murine aspartyl protease 2 (a) (Asp2) amino acid sequence.

Aspartyl protease: aspartate 1,1'-

aspartyl protease; aspartase; amyloid precursor protein; APP; Asp 2;
Alzheimer's disease; beta secretase site; neuro

XX beta secretase site; mouse.

Mus musculus.

AA
PN
WO200017369-A2.

XX

PD 30-MAR-2000.

1

QY 481 ADDISILK 488
 DB 494 addisilk 501

RESULT 12
 AAY88426
 ID AAY88426 standard; Protein; 476 AA.
 AC AAY88426;
 XX
 DT 03-AUG-2000 (first entry)
 XX
 DE Human aspartyl protease 2 (b) (Asp2) amino acid sequence.
 XX
 KW Aspartyl protease; aspartase; amyloid precursor protein; APP; Asp 2;
 KW Alzheimer's disease; beta secretase site.
 XX
 OS Homo sapiens.
 XX
 PN W0200017369-A2.
 XX
 PD 30-MAR-2000.
 XX
 PF 23-SEP-1999; 99WO-US20881.
 XX
 PR 24-SEP-1998; 98US-0101594.
 XX
 PA (PHAA) PHARMACIA & UPJOHN CO.
 XX
 PI Gurney ME, Blenkowski MJ, Heinrikson RL, Parodi LA, Yan R;
 XX
 DR WPI; 2000-303209/26.
 XX
 RX N-PSDB; AAA15663.
 XX
 PT New enzyme designated human aspartase useful in research into
 PT Alzheimer's Disease is capable of cleaving amyloid protein precursor at
 PT the beta secretase site to produce amyloid beta peptide -
 XX
 PS Claim 51; Fig 3; 183pp; English.
 XX
 CC This sequence represents the human aspartyl protease 2 (Asp2) amino acid
 CC sequence. The invention relates to a protease (e.g. Asp2) capable of
 CC cleaving the beta secretase site of amyloid precursor protein (APP). The
 CC protease contains a sequence encoding the amino acid sequence DTG and a
 CC sequence encoding DSG or DTG separated by 100-300 amino acids. When
 CC mutated the APP gene causes an autosomal dominant form of Alzheimer's
 CC disease. APP localises to the cell surface membrane and have a single
 CC C-terminal transmembrane domain. Proteolytic processing of APP produces
 CC the amyloid beta protein, which is possibly very important in Alzheimer's
 CC disease. The invention includes a nucleotide sequence encoding the
 CC protease, a vector containing the nucleotide sequence, and a cell line
 CC comprising the vector. Methods for screening for inhibitors of beta
 CC secretase activity are also given in the invention. The human aspartase
 CC protein and nucleotide sequences and the methods for identifying
 CC inhibitors of the protease, are useful in the treatment of and research
 CC in to Alzheimer's disease.
 XX
 SX Sequence 476 AA;

Query Match 94.0%; Score 2432.5; DB 21; Length 476;
 Best Local Similarity 94.9%; Pred. No. 3.3e-241;
 Matches 463; Conservative 0; Mismatches 0; Indels 25; Gaps 1;

QY 1 AGVLPAGHTGIRLPURSLGAGPLGLRLPRTDEPEEPGRGSGFVEMVNLRGKSGQ 60
 DB 14 agvlpaghtgqirlpurslgagplglrlpretdepeepgrgsgfvmvnlrgksqg 73
 QY 61 GYVEMTVGSPPTNLNLVDVTGSSNFAVGAAPHPFLHRYVQRLSSYRDLRGVYVPYT 120
 DB 74 gyyvemtvgspptnlilvdtgssnfavgaaphflhryyqrlsstyrdlrgvyvpyt 133

QY 121 QGKWEGLGTDLVSIPIPHGNVTVRANIAAITESDKFFINGSNWEGILGLAYAEIARPDSS 180
 DB 134 qgkwegeLgtdlvsiPhgnvtvrAniaaitesdkffingsnwegilglayaeIar---- 189
 QY 181 LEFFFDLSLVKQTHVPLNLSLQCGAGFPLNOSSEVLASVSGSMIIGIDHSLYTGSLWYTP 240
 DB 190 -----lcgagfplnqsevlasvsgsmilggldhslYtgsIwYtp 228
 QY 241 IRREWYVEIIVRVEINGODLKMDCKEYNYDKSIVDSGTTNLRPKKVFEEAAVKSIAAS 300
 DB 229 irrewyveIivrveingqdlkmdckeyndksivdsGttnlrlpkkvfeaaVksikaas 288
 QY 301 STEKFPDGFWLGEOLVCWQAGTTPWNIFFVISYLMGEVTNQSFRTITLPQQYLRPVEDV 360
 DB 289 stekfpdgfwlgeqlvcwqagttPwniffvislylMgevtngsfrtitlpqqYlrpvedv 348
 QY 361 ATSQDDCYKFAISOSSTGTVMGAVIMEGFYVDFDARKRIGFAVSACHVHDEFRTAAVEG 420
 DB 349 atsqqddcykfaisqsatgtvmgavimegfYvdfdrkrIgfavsachvhdefrtaaveg 408
 QY 421 PFVTLNEDCGYNIPQTDSTLMTIAYVMAAICALFMLPLCLMVCOWRCRLRQHQHDDF 480
 DB 409 pfvtlmedcgynipqtdestlmtIayvmaaaicalfmlplclmvcqwrcrlrqhqhdff 468
 QY 481 ADDISILK 488
 DB 469 addisilk 476

RESULT 13
 AAB07897
 ID AAB07897 standard; Protein; 456 AA.
 XX
 AC AAB07897;
 XX
 DT 14-NOV-2000 (first entry)
 XX
 DE Active enzyme portion of human beta-secretase enzyme.
 XX
 KW Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;
 KW amyloid plaque component; Alzheimer's disease; amyloidogenic disease;
 KW inhibitor; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200047618-A2.
 PD 17-AUG-2000.
 XX
 PF 10-FEB-2000; 2000WO-US03819.
 XX
 PR 10-FEB-1995; 99US-0119571.
 PR 15-JUN-1995; 99US-0139172.
 XX
 PA (ELAN-) ELAN PHARM INC.
 XX
 PI Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;
 PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;
 XX
 XX WPI: 2000-533011/48.
 XX
 PT Purified beta-secretase protein used in assays to discover inhibitors
 PT which can be used for the treatment of amyloidogenic diseases e.g.
 PT Alzheimer's disease -
 XX
 PS Claim 24; Fig 2B; 121pp; English.
 XX
 CC The specification describes a beta-secretase enzyme. The enzyme cleaves
 CC beta-amyloid precursor protein to produce beta-amyloid peptide. This
 CC enzyme is therefore implicated in the production of amyloid plaque
 CC components which accumulate in the brains of individuals afflicted with
 CC Alzheimer's disease. Inhibitors of beta-secretase are administered to

CC a mammalian subject e.g. with Alzheimer's disease or Alzheimer's
CC disease-like pathology to test if they maintain or improve cognitive
CC ability or reduce the plaque burden. The compounds are used for the
CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The
CC present sequence represents the active enzyme portion of human
CC beta-secretase enzyme.
XX Sequence 456 AA;
SQ

Query Match 93.5%; Score 2419; DB 21; Length 456;
Best Local Similarity 100.0%; Pred. No. 7.5e-240;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 33 ETDEEPEEPCRGSGFVEMVDNLGRKSGGGYVEMTVGSPQTLNLTVDGSSNPAVGAAP 92
Db 1 etdeepeepgrgsgfemvndnlrgksggggyvemtvgspqtlnlvtgssnfavgaap 60
QY 93 HPFLHRYQRLSSYRDLRGVYVYTGQKWEGLGTLVSIPIHGPNTVVRANIAATE 152
Db 61 hpfhryyqrqlssyrdlrgkgyvvytgqkwegeigtldvsihpgpntvvranaaite 120
QY 153 SKKFFINGNWEGILGLAYAEIARDDSLPEFFDLSLVKQTHVPLNLSLQLCGAGFPLNQS 212
Db 121 skkffingnwegilglayaeiarppdslepfldslvkqthvplnlsfqlcagagfplnqs 180
QY 213 EYLASVGGSMITGGIDHSLYGLSWYTPIRREWYEVITIVRVEINGDLMCKCKEYNYDK 272
Db 181 ewlasvvggsmi199idhslsglywtptirrewyevitivrveingdldmckckeynydk 240
QY 273 STVDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGLWGLVQWAGTTPWNIFFVIS 332
Db 241 sldvsgttnlrpkkvfeaaavksiaasstekfpdgfwlgeqlvcwaggttwnlffvis 300
QY 333 LYLMEVNTNQSFRTILPQQYLRPVEDVATSDQDCYKFAISQSSTGTVMGAVIMEGFYVV 392
Db 301 lylmevntnqsfritilppqylrpvedvatsdqdcykfaissqstgtvmgavimegyfvv 360
QY 393 FDRARKRIGFAVSACHVHDEPRTAAVEGPFVTLDMEDCGYNIPQDSTLMTIAYVMAAI 452
Db 361 fdrarkrigfavsachvhdefrtaavegpfvtdmedcgyinipqdestlmtiayvmaai 420
QY 453 CALFMLEPLCMVCWRCLRCRLRQHDHDFADDISLLK 488
Db 421 calfmplcmvcwrclrcrlrqhdhdfaddisllk 456

RESULT 14
ID AAY88438 standard; Protein; 453 AA.
XX AAY88438;
AC AAY88438;
XX
DT 03-AUG-2000 (first entry)
XX
DE Modified human aspartyl protease 2 (Asp2) amino acid sequence.
XX
KW Aspartyl protease; aspartase; amyloid precursor protein; APP; Asp 2;
KW Alzheimer's disease; beta secretase site.
XX
OS Homo sapiens.
XX
PN W0200017369-A2.
XX
PD 30-MAR-2000.
XX
PF 23-SEP-1999; 99WO-US200881.
XX
PR 24-SEP-1998; 98US-0101594.
XX
XX (PHAA) PHARMACIA & UPJOHN CO.
XX
XX Gurney ME, Blenkowski MJ, Heinrichson RL, Parodi LA, Yan R;
PI

XX WPI; 2000-303209/26.
DR N-PSDB; AAA15688.
XX
PT New enzyme designated human aspartase useful in research into
PT Alzheimer's disease is capable of cleaving amyloid protein precursor at
PT the beta secretase site to produce amyloid beta peptide -
XX
PS Example 10; Page 169-172; 183pp; English.
XX
XX This sequence represents a modified human aspartyl protease 2 (Asp2)
CC amino acid sequence. Asp2 encoded by this sequence has the C-terminal
CC transmembrane domain deleted. The invention relates to a protease
CC (e.g. Asp2) capable of cleaving the beta secretase site of amyloid
CC precursor protein (APP). The protease contains a sequence encoding the
CC amino acid sequence DTG and a sequence encoding a sequence encoding the
CC 100-300 amino acids. When mutated the APP gene causes an autosomal
CC dominant form of Alzheimer's disease. APP localises to the cell surface
CC membrane and have a single C-terminal transmembrane domain. Proteolytic
CC processing of APP produces the amyloid beta protein, which is possibly
CC very important in Alzheimer's disease. The invention includes a
CC nucleotide sequence encoding the protease, a vector containing the
CC screening for inhibitors of beta secretase activity are also given in the
CC invention. The human aspartase protein and nucleotide sequences and the
CC methods for identifying inhibitors of the protease, are useful in the
CC treatment of and research in to Alzheimer's disease.
XX
SQ Sequence 453 AA;

Query Match 90.0%; Score 2328; DB 21; Length 453;
Best Local Similarity 100.0%; Pred. No. 1.7e-230;
Matches 440; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGVLPAGHTQHGIRLPLRSGLGGAPLGLRLPRETDEEPEEGRGGSFVEMVDNLGRKSGQ 60
Db 14 agvlpaghtqhgirplrsglggapglrlpretdeeepeegrgrgsfemvndnlrgksqg 73
QY 61 GYVEMTVGSPQTLNLTVDGSSNPAVGAAPHPFLHRYQRLSSYRDLRGVYVYPT 120
Db 74 gyyvemtvgspqtlnlvtgssnfavgaaphflhryyqrqlssyrdlrgkgyvpyt 133
QY 121 QKWEGLGTLVSIPIHGPNTVVRANIAATESDKFFINGSNWEGILGLAYAEIARPPDS 180
Db 134 qkwegeigtldvsihpgpntvvranaiaatesdkffingsnwegilglayaeiarppds 193
QY 181 LEPPFDSLVKQTHVPLNLSLQLCGAGFPLNQSSEVLASVGGSMITGGIDHSLYGLSWYTP 240
Db 194 lepffdsilvkqthvplnlsfqlcagagfplnqsevlavsggsmi199idhslsglywtpt 253
QY 241 IRREWYEVITIVRVEINGDLMCKCKEYNYDKSIVDSGTTNLRPKKVFEEAAVKSIAAS 300
Db 254 irrewyevitivrveingdldmckckeynydksivdsgettlnlrpkkvfeaaavksikaas 313
QY 301 STEKFPDGLWGLVQWAGTTPWNIFFVISLYLMEVNTNQSFRTILPQQYLRPVEDV 360
Db 314 stekfpdgfwlgeqlvcwaggttwnlffvislylmevntnqsfritilppqylrpvedv 373
QY 361 ATSQDDCYKFAISQSSTGTVMGAVIMEGFYVYVDFDARKRIGFAVSACHVHDEFRFAAVEG 420
Db 374 atsqqdcykfaissqstgtvmgavimegyfvvdfdrarkrigfavsachvhdefrfaaveg 433
QY 421 PFVTLDMEDCGYNIPQDSTDES 440
Db 434 pfvtldmedcgyinipqtdes 453

RESULT 15
ID AAY88439 standard; Protein; 459 AA.
XX AAY88439;
AC AAY88439;

XX 03-AUG-2000 (first entry)
DT Modified human aspartyl protease 2 (Asp2) amino acid sequence.
XX
DE Aspartyl protease; aspartase; amyloid precursor protein; APP; Asp 2;
KW Alzheimer's disease; beta secretase site.
KW
XX Homo sapiens.
OS
XX WO200017369-A2.
XX
XX 30-MAR-2000.
XX
XX 23-SEP-1999; 99WO-US20881.
XX
XX 24-SEP-1998; 98US-0101594.
XX
XX (PHAA) PHARMACIA & UPJOHN CO.
XX
XX Gurney ME, Blenkowski MJ, Heinrichson RL, Parodi LA, Yan R;
PI
XX WPI; 2000-303209/26.
DR
DR N-PSDB: AAA15689.
XX
XX New enzyme designated human aspartase useful in research into
PT Alzheimer's Disease is capable of cleaving amyloid protein precursor at
PT the beta secretase site to produce amyloid beta peptide -
XX
XX Example 10; Page 173-176; 183pp; English.
XX
XX This sequence represents a modified human aspartyl protease 2 (Asp2)
CC amino acid sequence. Asp2 encoded by this sequence has the C-terminal
CC transmembrane domain deleted. The invention relates to a protease
CC (e.g. Asp2) capable of cleaving the beta secretase site of amyloid
CC precursor protein (APP). The protease contains a sequence encoding the
CC amino acid sequence DTG and a sequence encoding DSG or DTG separated by
CC 100-300 amino acids. When mutated the APP gene causes an autosomal
CC dominant form of Alzheimer's disease. APP localises to the cell surface
CC membrane and have a single C-terminal transmembrane domain. Proteolytic
CC processing of APP produces the amyloid beta protein, which is possibly
CC very important in Alzheimer's disease. The invention includes a
CC nucleotide sequence encoding the protease, a vector containing the
CC nucleotide sequence, and a cell line comprising the vector. Methods for
CC screening for inhibitors of beta secretase activity are also given in the
CC invention. The human aspartase protein and nucleotide sequences and the
CC methods for identifying inhibitors of the protease, are useful in the
CC treatment of and research in to Alzheimer's disease.
XX
XX Sequence 459 AA;
SQ

Query Match 90.0%; Score 2328; DB 21; Length 459;
Best Local Similarity 100.0%; Pred. No. 1.7e-230;
Matches 440; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGVLPAGHTQHGIRLPLRSLGGAPLGLRLPRETDEEPEGRGSGFVEMVDNLRKSGQ 60
Db 14 agvlpaghtqhgirplrsllggaplglrlpretdeepegrgsgfvmvndnrlrksqg 73
QY 61 GYVEMTVGSPQDTNLNLTVDGSSNFAGVGAAPHPFLHRYQRLSSTYRDLRKGVIYPT 120
Db 74 gyyvemtvgspqdtlnlilvtgssnfavgaaphflhryyqrlsstyrdlrkgviypt 133
QY 121 QGKWEGLGTDLYSIPHPNVTVRANIAATSDKFFINGSNWEGILGLAYAEIARPDSS 180
Db 134 qgkweglgtdlysihpgnvtvranaiaatesdkffingsnwegilglayaeiarpdds 193
QY 181 LEPPFDSLVKQTHVPNLFSLQLCCAGFPPLNQSSEVLASVSGSMIIGGIDHSLYTGSWYTP 240
Db 194 lepfdsylvkqthvpnlfsllqlcagfplnqsevlasvsgsmiiggidhsltygslwtp 253
QY 241 IRREWYEVIIIVRVEINGQDLKMDCKEYNDKSIDVSGTTLNLRPKKVFEEAVKSIKAAS 300

Db 254 irrewyeviiivrveingqdlkmdckeyndksivdsgttnlrpkkvfeaaavksikaas 313
QY 301 STEKFPDGFGLGQVLCWQAGTTPWNIFPVISLYLMGEVTNQSFRTITILPQOYLRPVEDV 360
Db 314 stekfpdgfwlgeqlvcwagttppwnifpvvislylmgevtngsfritilpqqlrpvedv 373
QY 361 ATSODDCYKFAISQSSTGTVMGAVIMEGFYVVPDRARKRIGFAVSACHVHDEPRTAAVEG 420
Db 374 atsqddcykfalsqsstgtvmgavimegfyyvfdrrarkrigfavsachvhdeirtaaveg 433
QY 421 PFVTLDMEDCGYNIPOTDES 440
Db 434 pfvldmedcgynipqtcdes 453

Search completed: September 6, 2001, 16:43:25
Job time: 354 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:37:31 ; Search time 113.12 Seconds
(without alignments)
88.827 Million cell updates/sec

Title: US-09-603-713-2
Perfect score: 2587
Sequence: 1 AGVLPANGTQHIGIRLPLRSG.....CLRCRQHQHDFADDISLLK 488

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2.6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2.6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2.6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2.6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2.6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2.6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1175	45.4	518	3	US-08-999-723-2
2	1175	45.4	518	4	US-09-434-427-2
3	299.5	11.6	396	1	US-08-208-007A-13
4	299.5	11.6	396	4	US-09-032-523-9
5	296.5	11.5	412	1	US-08-208-007A-12
6	296.5	11.5	412	4	US-08-974-691-4
7	279.5	10.8	458	6	5217891-15
8	273.5	10.6	409	1	US-08-360-673-6
9	273.5	10.6	427	2	US-08-846-021A-8
10	270	10.4	410	1	US-08-088-633-2
11	270	10.4	410	1	US-08-245-756-2
12	270	10.4	410	1	US-08-441-750-2
13	270	10.4	410	2	US-08-441-751-2
14	270	10.4	410	5	PCT-US92-02521-2
15	250	9.7	349	4	US-09-032-523-3
16	232	9.0	398	1	US-08-328-314-2
17	232	9.0	398	1	US-08-731-045-2
18	216.5	8.3	397	3	US-09-079-415-2
19	215	8.3	419	4	US-08-974-691-3
20	203	7.8	427	1	US-07-958-222A-2
21	197	7.6	419	3	US-08-115-753-2
22	197	7.6	419	3	US-08-115-753-33
23	196.5	7.6	420	4	US-09-008-271A-4
24	196.5	7.6	420	4	US-08-974-691-8
25	196.5	7.6	430	1	US-08-535-237-2
26	194	7.5	395	1	US-08-723-938-3
27	194	7.5	395	2	US-09-080-538-3

28 193 7.5 330.3 3 US-08-115-753-1 Sequence 1, Appl1
29 193 7.5 445 4 US-08-974-691-6 Sequence 6, Appl1
30 193 7.5 451 4 US-08-974-691-2 Sequence 2, Appl1
31 129.5 5.0 140 3 US-09-211-631-13 Sequence 13, Appl1
32 129.5 5.0 140 4 US-09-265-628-13 Sequence 13, Appl1
33 129.5 5.0 140 4 US-09-001-141-11 Sequence 11, Appl1
34 97 3.7 1030 4 US-09-091-117-2 Sequence 2, Appl1
35 95.5 3.7 280 4 US-09-160-246-14 Sequence 14, Appl1
36 92.5 3.6 1097 2 US-08-680-326-39 Sequence 39, Appl1
37 88 3.4 746 2 US-08-838-219B-6 Sequence 6, Appl1
38 88 3.4 746 3 US-09-233-336A-6 Sequence 6, Appl1
39 88 3.4 746 4 US-09-233-752A-6 Sequence 6, Appl1
40 88 3.4 789 1 US-08-471-033-32 Sequence 32, Appl1
41 88 3.4 789 2 US-08-471-044-32 Sequence 32, Appl1
42 88 3.4 789 2 US-08-463-483A-32 Sequence 32, Appl1
43 88 3.4 789 2 US-08-471-046A-32 Sequence 32, Appl1
44 88 3.4 789 2 US-08-470-566B-32 Sequence 32, Appl1
45 88 3.4 789 2 US-08-838-219B-4 Sequence 4, Appl1

ALIGNMENTS

RESULT 1
US-08-999-723-2
; Sequence 2, Application US/08999723A
; Patent No. 6035180
; GENERAL INFORMATION:
; APPLICANT: Powell, David J.
; APPLICANT: Southan, Christopher
; APPLICANT: Chapman, Conrad G.
; APPLICANT: Evans, Joanne R.
; TITLE OF INVENTION: ASPI
; FILE REFERENCE: GH70262
; CURRENT APPLICATION NUMBER: US/08/999,723A
; CURRENT FILING DATE: 1997-10-06
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-999-723-2

Query Match 45.4%; Score 1175; DB 3; Length 518;

Best Local Similarity 47.1%; Pred. No. 1.5e-117;

Matches 228; Conservative 81; Mismatches 151; Indels 24; Gaps 5;

QY 5 PAHGTQHIGIRLPLRSGGGAPLGLRLPRETDEPEEPGPRGSGFVEMVDNLRGKSGGYV 64

DB 52 PAERHADGLALALEPALA-----SPAGAAFLAMVDNLQDGSRGYYL 94

QY 65 EMVVGSPQPLNLILVDVGSSNEAVGAAPHPLHRYQORLSSTYRDLRKGVVYPYTGKW 124

DB 95 EMLGTPPQKLQILDVDFGGSNFVAVAGTPHSDYIDFTFSTSYRSKGFDTVYKTCGW 154

QY 125 EGELGTDLVSIPIHGPVNTVRANITAAITESDKFFINGSNWEGILGLAYAEIARPDSDLPEF 184

DB 155 TGFVGDVLTPIKGFNTSLVNIATIFESFNFFLPKIKWNGILGLAYATLAKPSSLET 214

QY 185 FDSLVRQTHVPLNLFSLQCGAPPLNOSVLA SVGSGMIIGGDHSLYTSGLWYTPTRRE 244

DB 215 FDSLVTQANIPNVFSGMCGAGLPVAGS---GTNGSGSLVGLGIEPSLYKGDINWYTPKEE 271

QY 245 WYEVIVTVRYEINGODLKMCCKEYNDKSIYDSCITNLRPLPKVFEAAVKSIAASSTEK 304

DB 272 WYVOIEFLKLEIGQSUNLDCREYNADKALVDGTTLLRLPQKQFVDAVAVARASLIPE 331

QY 305 FPDGFWLGEOLVCWQAGCTTPWNIPFPVISLYLMGEVNTNOSFRITILPOOYLRPVEDVATSQ 364

DB 332 FSDGFWTGSQACWNTSETWSPFKISIVLRDENSRSFRITILPOLYIQPMWAGLNY 391

[illegible]

Db 276 COAIVDTGSLTGPDSKIKQLQNAIGAAP-----VDGEYAVE-----CANLNVMPD 322
QY 331 ISLYLMGEVNTQSFRTITLPOQYLRVEDVATSQDDCYKFAISQSTG----- 378
Db 323 VFTING-----VPYTLSTAY--TLLDFVDMQFC-----SSGFGGLDIHPPAGP 366

QY 379 -TVMGAVIMEGFYVDFDRARRKIGFA 403
Db 367 LWILGDVFIRQFYVDFDRGNRRVGLA 392

RESULT 4
US-09-032-523-9
; Sequence 9, Application US/09032523
; Patent No. 6232454
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl
; APPLICANT: Baugh, Mariah
; TITLE OF INVENTION: HUMAN PROTEINASE MOLECULES
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
FILING DATE:
APPLICATION NUMBER: US/09/032,523
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0479 US
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:

INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 396 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Genbank
CLONE: 181994
US-09-032-523-9

Query Match 11.6%; Score 299.5; DB 4; Length 396;
Best Local Similarity 25.9%; Pred. No. 1.5e-23;
Matches 100; Conservative 65; Mismatches 148; Indels 73; Gaps 16;

QY 35 DEEPPEPGRGSGFVEMVDNLRGSGGQYVEMTVGSPQTLNLVDTGSSNFVGA---- 90
Db 63 DQSAKEP-----LNYLD-----MEYFTIGSGPPQNFTVIFDTGSSNLWVPSVYCT 110
QY 91 APHPFLHRYQRLSTYDLRKGVVVPTQGWKEGLCTDLVSIHPGPNVTVRANAAI 150

Db 111 SPACKTHSRFQPSQSTYSQPCQSFSIQYTGSLSGIIGADQSV--EGLTVVQOQGSEV 169
QY 151 TESDRFFINGSNWEGILGAYAEIARPDSDSLRPFDDSLVKQTHVPNLFSLQLCGAGFPLN 210
Db 170 TEPGQTFVD-AEFDGLGLGYPSLA--VGGVTPVFDNMAQ-----NLVDLPMFVSVMSSN 222
QY 211 QSEVLASVCGSMIIGDHSHTGSLWYTPIRREWYEVYIIVRVEINGODLKMDCHEYNY 270
Db 223 PE---GGAGSELIFGGYDHSFSGSLNMYVPTKQAYWQIALDNIQVGG--TVMFCSE--G 275
QY 271 DKSIVDSGTTNLRPKKVFEEAAVKSIKAASSTKPKDPDFWLGEOLYCVMQAGTTPWNIFFV 330
Db 276 COAIVDTGSLTGPDSKIKQLQNAIGAAP-----VDGEYAVE-----CANLNVMPD 322
QY 331 ISLYLMGEVNTQSFRTITLPOQYLRVEDVATSQDDCYKFAISQSTG----- 378
Db 323 VFTING-----VPYTLSTAY--TLLDFVDMQFC-----SSGFGGLDIHPPAGP 366
QY 379 -TVMGAVIMEGFYVDFDRARRKIGFA 403
Db 367 LWILGDVFIRQFYVDFDRGNRRVGLA 392

RESULT 5
US-08-208-007A-12
; Sequence 12, Application US/08208007A
; Patent No. 5501969
; GENERAL INFORMATION:
; APPLICANT: HASTINGS, ET AL.
; TITLE OF INVENTION: Human Osteoclast-Derived Cathepsin
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSES: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSES: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/208,007A
FILING DATE: March 8, 1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: No. 5501969e
FILING DATE: No. 5501969e
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-95
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 412 AMINO ACIDS
TYPE: AMINO ACID
STRANDEDNESS:
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
US-08-208-007A-12

Query Match 11.5%; Score 296.5; DB 1; Length 412;
Best Local Similarity 26.8%; Pred. No. 3.3e-23;
Matches 117; Conservative 66; Mismatches 163; Indels 91; Gaps 19;
QY 13 IRLPLR-----SGLGAPLGL-----RLPRETDEEPEEPGRGGSFVEMVD 52

Db 152 YFGIYGLTPPOEFTVLDFTGSSDFW---PSIYCKSNACKNHQDQRKSSFTQNLGKDL 208
 QY 116 VVPYTGKWEGLGDLVSIHPGNVTVRANAAITSDKFFINGSNWEGLGLAYAEIA 175
 Db 209 SHYGTGSMQGLGDTVTNINVIDIQTVGLSTGEPODF--TFAEFGIILOMAYPSUA 266
 QY 176 RPDSLEPFDSLVRQTHV-PNLSIQLCGAGFPPLNQSEVLASVSGSMITIGIDHSLYTG 234
 Db 267 --SEYSPVFDNMNRHLVAQDLFSVYMDRNG---QESMLT-----LGAIDPSYITG 313
 QY 235 SLWYPIREWEYEVIIIVRVEINGDGLKMD--CKEYNYDKSIVDSCTNLRPKKVFEEA 292
 Db 314 SLHWVPTVQOQWQFTVDSVTSISGVVACEGGC-----QALDGTSLKLVGSPSSDILMI 367
 QY 293 VKSIAASSTERFPGDFWGLGE-QLVCWQAGTTPWNIFFVISLYLMGEVNTQFRITILPQ 351
 Db 368 QAIGATQNG-----YGEFDIDCNLSYMTVVF-----EINGKMYPLT--PS 408
 QY 352 QYLRPVEDVATSDDCY---KFAISOSSTGTVMGAVIMEGFYVVDRAKRIGFA 403
 Db 409 AY-----TSQDQGFCTSGFQSNHSGQWILGDVFIREYYSVFDRAANLVGLA 455

RESULT 8
 US-08-360-673-6
 ; Sequence 6, Application US/08360673
 ; Patent No. 5679544
 ; GENERAL INFORMATION:
 ; APPLICANT: Fleer, Reinhard
 ; APPLICANT: Fournier, Alain
 ; APPLICANT: Yeh, Patrice
 ; TITLE OF INVENTION: MODIFIED KLUYVEROMYCES YEASTS, THEIR
 ; NUMBER OF SEQUENCES: 17
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Rhone-Poulenc Rorer Inc.
 ; STREET: 500 Arcola Rd. 3C43
 ; CITY: Collegeville
 ; STATE: PA
 ; COUNTRY: USA
 ; ZIP: 19002
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/360,673
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: WO PCT/FR93/00623
 ; FILING DATE: 23-JUN-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: FR 92/07785
 ; FILING DATE: 25-JUN-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Smith, Julie K.
 ; REGISTRATION NUMBER: 38,619
 ; REFERENCE/DOCKET NUMBER: ST92040-US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (610)454-3839
 ; TELEFAX: (610)454-3808
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 409 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-360-673-6

Query Match 10.6%; Score 273.5; DB 1; Length 409;

Best Local Similarity 25.3%; Pred. No. 9.8e-21;
 Matches 92; Conservative 59; Mismatches 144; Indels 69; Gaps 11:
 QY 62 YVVENTGSGPPOTLNLIVDTGSSNFAVGAAP---HPFLHRYVYQROLSSYRDLRKGVV 117
 Db 96 YFTEITLGGSPPOQSFVKVILDTGSSNLWVPSAEGCSLACFLHTYDHEASTYKANGSEPAI 155
 QY 118 PYTOGKWEGLGDLVSIHPGNVTVRANAAITSDKFFINGSNWEGLGLAYAEIARP 177
 Db 156 QYSGSLEGVSRDLTI--GDLVLPDQDFAEATSEPLAFAGFKFDGILGLAYDSIS-V 212
 QY 178 DDSLPEPFDLSLVKQTHVFNLSIQLCGAGFPPLNQSEVLASVSGSMILGIDHSLYTSLW 237
 Db 213 NRIVPPVYNAIKNLLDDPVFA-----FYLGDSD--KSEDDGEASFGGIDEEKYTGTEIT 264
 QY 238 YPIRREWEYEVIIIVRVEINGDGLKMDCKEYNYDKSIVDSCTNLRPKKVFEEAVKSIK 297
 Db 265 WLPVRKAWE-----VKFEGIGLGEYATLEHGGAIDTGLSLIALPSGLAEILNAEIG 319
 QY 298 AASSTKFPDGFWLQVLCWQAGTTPWNIFFVISLYLMGEVNTQSFRT----- 347
 Db 320 AKKG-----WSQYSVDCESRDS---LPDLTLNFNG---YNTTITAYDTTLEVSG 363
 QY 348 -----ILPOQLRPVEDVATSDDCYKFAISQSSTGTVMGAVIMEGFYVVDRAKRIGF 402
 Db 364 SCISAFTPMDFPEPVGFLA-----IIGDAFLRKYYSIYDIGHDAVGL 405
 QY 403 AVSA 406
 Db 406 AKAA 409
 RESULT 9
 US-08-846-021A-3
 ; Sequence 8, Application US/08846021A
 ; Patent No. 5948682
 ; GENERAL INFORMATION:
 ; APPLICANT: Moloney, Maurice M.
 ; TITLE OF INVENTION: Preparation of Heterologous Proteins on
 ; TITLE OF INVENTION: Oil Bodies
 ; NUMBER OF SEQUENCES: 31
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: BERESKIN & PARR
 ; STREET: 40 King Street West
 ; CITY: Toronto
 ; STATE: Ontario
 ; COUNTRY: Canada
 ; ZIP: M5H 3Y2
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/846,021A
 ; FILING DATE: April 25, 1997
 ; CLASSIFICATION: 800
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Gravelle, Micheline
 ; REGISTRATION NUMBER: 40,261
 ; REFERENCE/DOCKET NUMBER: 9369-039
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (416) 364-7311
 ; TELEFAX: (416) 361-1398
 ; INFORMATION FOR SEQ ID NO: 8:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 427 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-846-021A-8

```
Query Match      10.6%; Score 273.5; DB 2: Length 427;
Best Local Similarity 26.4%; Pred. No. 1.1e-20;
Matches 94; Conservative 68; Mismatches 129; Indels 65; Gaps 15;

QY 62 YVEMTVGSPQTNLNLTGSSNFVAGAPHFL-----HRYQRLSSTYRDLRKG 114
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 120 YFGKIYLGTPQEFVLTGSSDEWV---PSIYCKSNACKNHRDFPKSSTFQNLGKP 176
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 115 VVYPYTOGWKEGELGDLVSIHPGPNVTVRANIAITSDKFFINGSWEGILGLAYAEI 174
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 177 LSIHYGTSMOGLIGYDVTIVSN--IVDIQQTVGLSTPEPGDVFYAEFDGLGMAYPSL 234
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 175 ARDDSLSEFFEDSLVKQTHV--PNLFSLOLCGAGFPLNQSEVLASVGGSMIIGGIDHSLYT 233
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 235 A--SEYSIPVDNMNRHLVAODLESVMYDRNG-----QESMLT-----LCAIDPSYTT 281
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 234 GSLWYTPIRREWYEVIIIVRVEINGDQKMD--CKEYNYDKSIVDSGTTNLRPKKVFEEA 291
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 282 GSLHWVPVTVQQYMQFTVDSVTISGVVAVACEGC-----QAILDTGTSKLVGPSDDLIN 335
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 292 AVKSIKAASSTEFKPDGFWLGE--OLVCWQAGTTPWNIPFVISLYLMGEVTVNSFRITILP 350
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 336 IQQAIGATQNG-----YGEFIDICDNLNLYMTVVF-----EINGKMYPLT--P 376
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 351 QQYLRPVEDVATSDQDCY---KFAISQSSTGTVMGAVIMEGFVYVDFDRARRKIGFA 403
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 377 SAY-----TSQDQGFCTSGFQSENHSQKWILGDFVIREYYSVDFDRANLVLGA 424
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
US-08-088-633-2
; Sequence 2, Application US/08088633
; Patent No. 5324660
; GENERAL INFORMATION:
; APPLICANT: Gleeson, Martin A
; APPLICANT: Howard, Bradley D
; TITLE OF INVENTION: Genes which Influence Pichia proteolytic
; TITLE OF INVENTION: Activity, and Uses Therefor
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fitch, Even, Tabin & Flannery
; STREET: 135 South LaSalle Street, Suite 900
; CITY: Chicago
; STATE: Illinois
; COUNTRY: U.S.A.
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/088,633
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/678,916
; FILING DATE: 01-APR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E
; REGISTRATION NUMBER: 31192
; REFERENCE/DOCKET NUMBER: 50848
; TELEPHONE: (619)552-1311
; TELEFAX: (619)552-0095
; TELEX: 20 6566 PATLAW CGO
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 410 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-088-633-2
```

```
Query Match      10.4%; Score 270; DB 1: Length 410;
Best Local Similarity 25.7%; Pred. No. 2.3e-20;
Matches 91; Conservative 51; Mismatches 162; Indels 50; Gaps 10;

QY 62 YVEMTVGSPQTNLNLTGSSNFVAGV----APHPFLHRYQRLSSTYRDLRKGVVY 117
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 96 YFTEVSLGTPQSFVKVILDTGSSNLWVPKDCGSLACFLHAKYDHDESSTYKKNSSFEI 155
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 118 PYTGKWEGLGTDLVSIHPGPNVTVRANIAITSDKFFINGSWEGILGLAYAEIARP 177
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 156 RYSGSMEGYVSDVLOI--GDLTPKVDFAEATSEPLAFAFKFGDILGLAY----- 207
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 178 DLSLEPFDSLVKQTHVPNLF---SLQLCGA---GFPLNQSEVLASVGGSMIIGGIDHSL 231
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 208 -----DSISVANKIYPIYKALELDDLPKFAFYLGDFDKDESGLATFGVDKSK 259
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 232 YGSLWYTPIRREWYEVIIIVRVEINGDQKMDCKEYNYDKSIVDSGTTNLRPKKVFEEA 291
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 260 YEGKITWLPVRRKAYWE-----VSFDGVGLGSEVAELOKTCAAIDTGTSLIALPSGLAEI 314
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 292 AVKSIKAASSTEFKPDGFWLGEOLVCWQAGTTPWNIPFVISLYLMGEVTVNSFRITILPQ 351
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 315 LNAEIGATKG-----WSGOYAVDCDTRDS---LPDLTLTFAG-----YNFITIPY 356
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 352 OYLRPVEDVATSDQDCYKFAISQSSTGTVMGAVIMEGFVYVDFDRARRKIGFAVS 405
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 357 DYLEVSGSCISAFTPMDFP-EPICPLAIGDSFLRKYYSVVDLGKDAVGLAKS 409
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
US-08-245-756-2
; Sequence 2, Application US/08245756
; Patent No. 5541112
; GENERAL INFORMATION:
; APPLICANT: Gleeson, Martin A
; APPLICANT: Howard, Bradley D
; TITLE OF INVENTION: GENES WHICH INFLUENCE PICHIA PROTEOLYTIC
; TITLE OF INVENTION: ACTIVITY, AND USES THEREFOR
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PRETTY, SCHROEDER, BRUEGGEMANN & CLARK
; STREET: 444 So. Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/245,756
; FILING DATE: 16-May-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/088,633
; FILING DATE: 06-JULY-1993
; PRIOR APPLICATION NUMBER: 07/678,916
; FILING DATE: 01-APR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 9763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-546-4737
; TELEFAX: 619-546-9392
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 410 amino acids
```

;
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-245-756-2

Query Match 10.4%; Score 270; DB 1; Length 410;
Best Local Similarity 25.7%; Pred. No. 2.3e-20;
Matches 91; Conservative 51; Mismatches 162; Indels 50; Gaps 10;

QY 62 YVEMTVGSPPTLNTLVDTGSSNFAVGA---APHFHLHRYQROLSSYRDLRKGVV 117
DB 96 YTEVSLGTTPGOSFKVILDTGSSNLWVPKDCGSLACFLHAKYDHDESSTYKNGSSFEI 155
QY 118 PYTQKWEGLGTDLVSIPIHGPNNVTVRANIAAITESSDKFFINGSNWEGILGLAYAEIARP 177
DB 156 RYGSMEGVSDVLIQI--GDLTIKVDFAEATSEPLAFKFGDGLILGLAY----- 207
QY 178 DLSLEPFFDSLVKQTHVNLNLF---SLQLCGA---GPPNQSEVLASVGGSMIIGDHS 231
DB 208 -----DSISVKNKTVPPPIYKALELDLDEPKFAFYLGDTDKDESDGLATFGGV 259
QY 232 YTGSLWYTPIRREWEYEVIIIVRVEINGODLKMDCKEYNDKSIDVSDGTTNLRPLPKVFEA 291
DB 260 YEGKITWLPVRRKAYWE-----VSFGVGLGSEYAELOKTGAIDTGTSLIALPSGLAEI 314
QY 292 AVKSIKAASSTKFPDGFGLGQVLCVQAGTTPWNTFPVISLYLMGEVNTNQSFRITILPQ 351
DB 315 LNAEICATKG-----WSGOYAVDCDTRDS---LPDLTLTFAG-----YNFTITPY 356
QY 352 QYLRPVEDVATSDCCYKFAISOSSTGTVMGAVIMSGFYVVDPRKRIGFAVS 405
DB 357 DYTLEVSGSCISAFTPMDFP-EPIGPLAIIGDSFLRKYYSVYDLGKDAVGLAKS 409

RESULT 12

US-08-441-750-2
; Sequence 2, Application US/08441750
; Patent No. 5691166
; GENERAL INFORMATION:
; APPLICANT: Gleeson, Martin A
; APPLICANT: Howard, Bradley D
; TITLE OF INVENTION: GENES WHICH INFLUENCE PICHIA PROTEOLYTIC
; TITLE OF INVENTION: ACTIVITY, AND USES THEREFOR
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/441,750
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/245,756
; FILING DATE: 16-May-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/088,633
; FILING DATE: 06-JULY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/678,916
; FILING DATE: 01-APR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie

;
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 9763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-238-0999
; TELEFAX: 619-238-0062
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 410 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-441-750-2

Query Match 10.4%; Score 270; DB 1; Length 410;
Best Local Similarity 25.7%; Pred. No. 2.3e-20;
Matches 91; Conservative 51; Mismatches 162; Indels 50; Gaps 10;

QY 62 YVEMTVGSPPTLNTLVDTGSSNFAVGA---APHFHLHRYQROLSSYRDLRKGVV 117
DB 96 YTEVSLGTTPGOSFKVILDTGSSNLWVPKDCGSLACFLHAKYDHDESSTYKNGSSFEI 155
QY 118 PYTQKWEGLGTDLVSIPIHGPNNVTVRANIAAITESSDKFFINGSNWEGILGLAYAEIARP 177
DB 156 RYGSMEGVSDVLIQI--GDLTIKVDFAEATSEPLAFKFGDGLILGLAY----- 207
QY 178 DLSLEPFFDSLVKQTHVNLNLF---SLQLCGA---GPPNQSEVLASVGGSMIIGDHS 231
DB 208 -----DSISVKNKTVPPPIYKALELDLDEPKFAFYLGDTDKDESDGLATFGGV 259
QY 232 YTGSLWYTPIRREWEYEVIIIVRVEINGODLKMDCKEYNDKSIDVSDGTTNLRPLPKVFEA 291
DB 260 YEGKITWLPVRRKAYWE-----VSFGVGLGSEYAELOKTGAIDTGTSLIALPSGLAEI 314
QY 292 AVKSIKAASSTKFPDGFGLGQVLCVQAGTTPWNTFPVISLYLMGEVNTNQSFRITILPQ 351
DB 315 LNAEICATKG-----WSGOYAVDCDTRDS---LPDLTLTFAG-----YNFTITPY 356
QY 352 QYLRPVEDVATSDCCYKFAISOSSTGTVMGAVIMSGFYVVDPRKRIGFAVS 405
DB 357 DYTLEVSGSCISAFTPMDFP-EPIGPLAIIGDSFLRKYYSVYDLGKDAVGLAKS 409

RESULT 13

US-08-441-751-2
; Sequence 2, Application US/08441751
; Patent No. 5831053
; GENERAL INFORMATION:
; APPLICANT: Gleeson, Martin A
; APPLICANT: Howard, Bradley D
; TITLE OF INVENTION: GENES WHICH INFLUENCE PICHIA PROTEOLYTIC
; TITLE OF INVENTION: ACTIVITY, AND USES THEREFOR
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/441,751
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/245,756
; FILING DATE: 16-May-1994
; CLASSIFICATION: 435


```

; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/032,523
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0479 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 349 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PANCUT01
; CLONE: 1515165
;
US-09-032-523-3
```

```

Query Match          9.7%; Score 250; DB 4; Length 349;
Best Local Similarity 23.1%; Pred. No. 2.5e-18;
Matches 86; Conservative 57; Mismatches 136; Indels 94; Gaps 13;

QY 35 DEPEEPGRGSEVMVDNLGRKGOGYVEMTVGSPPOTLNLLVDTGSSNEAVGA---- 90
Db 63 DQSAKEP-----LINVLD-----MEYFTISIGPPQNFTVIFDGTSSNLWVPSVYCT 110

QY 91 APHPFLHRYQRLSSTYDLRKGVVYPTQGWEGELGDLVSIHPGPNVTVRANIAAI 150
Db 111 SPACKTHSRFPQSSTYSQPGSFSIQYGTGSLGIIGADQYSV-EGLTVVGQGFGEV 169

QY 151 TESDKFFINGSNNEGILGLAYAEIARPDSDLFPFDSLKVTQTHVPNLFSLQLCGAGPPLN 210
Db 170 TEPQOTFVD-AEFDGILGLGPSLA--VGGVTPVFONMMAQ-----NLVDLPMEFSVYMSSN 222

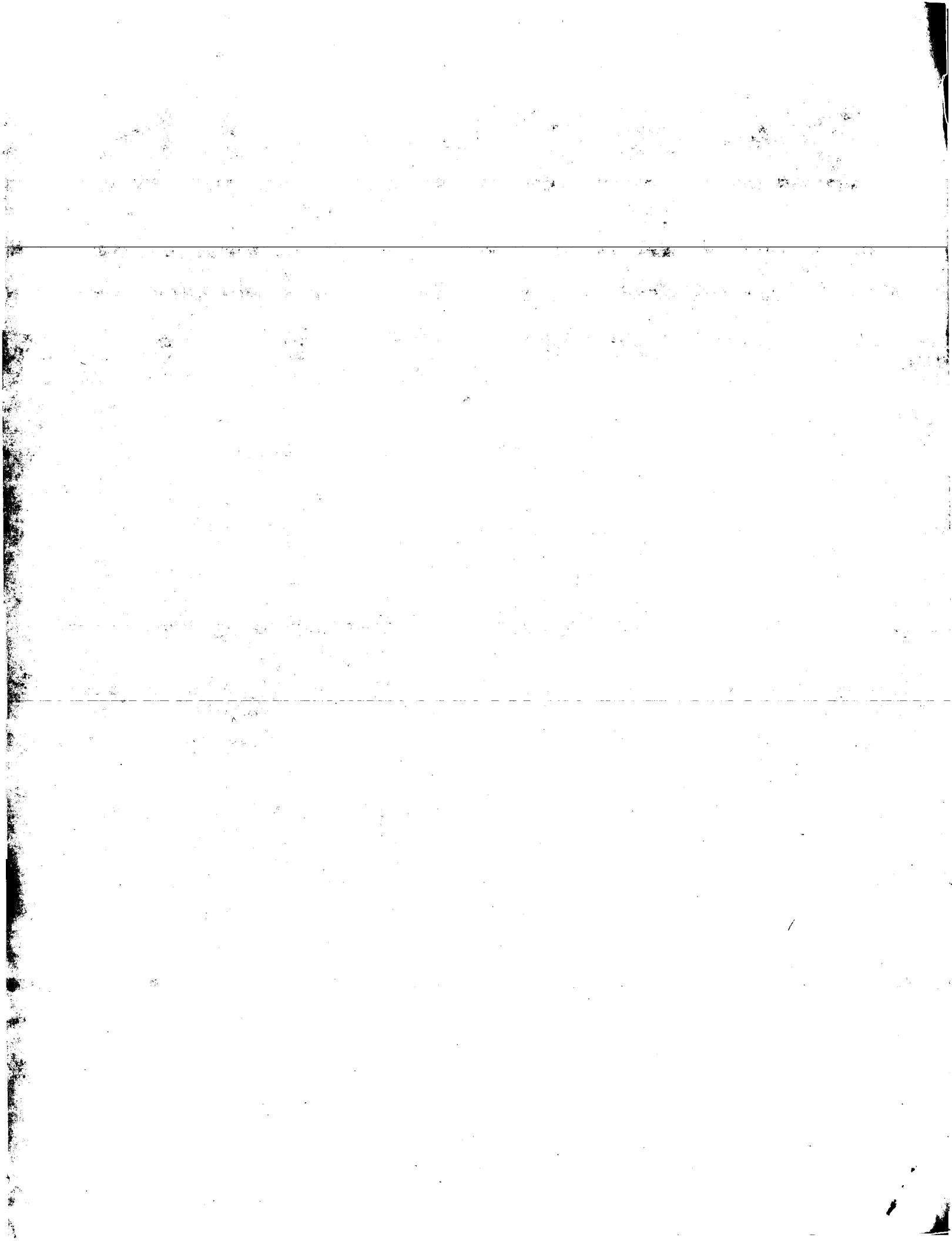
QY 211 QSEVLASVGGSMIIGIDHSLYTGSLWYTPIRREWYEVIIIVRVEINGQDLKMDCKEYNY 270
Db 223 PE--GGAGSELIFGGYDHSFSGSLNWPVTQKQAYWQIAL----- 260

QY 271 DKSIVDSGTTNLRPKRVFAAVKSIKAASSTKFPDGFMLGEQLVCWQAGTTPWNIFPV 330
Db 261 DNYAVECANLV-----MPD-----VTFTINGVVPYTLSP 290

QY 331 ISLYLMGEVTNQSFRTILPQQYLRPVEDVATSDCCYKFAISQSSTGTVMGAVIMEGFY 390
Db 291 -AYTLDFVDGMQFCSSGFGQLDHP-----PAGPLWILGDVFIQFY 332

QY 391 VVFDRAKRKIGFA 403
Db 333 SVFDGRGNRVGLA 345
```

Search completed: September 6, 2001, 16:39:33
Job time: 122 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:37:36 ; Search time 134.15 Seconds
(without alignments)
277.102 Million cell updates/sec

Title: US-09-603-713-2

Perfect score: 2587

Sequence: 1 AGVLPAGTQHGIRLPLRSG.....CLRLRQHQHDFADDISLLK 488

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_68:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2587	100.0	501	2 A59090	aspartic proteinase
2	314	12.1	384	3 JC7574	pepsinogen A - Afr
3	308	11.9	385	3 JC7575	pepsinogen A - bul
4	306	11.8	387	2 B38302	pepsin (EC 3.4.23)
5	304.5	11.8	383	3 JC7573	pepsinogen C - Afr
6	304.5	11.8	388	1 S19682	pepsin A (EC 3.4.2)
7	302	11.7	384	2 A39314	gastricsin (EC 3.4)
8	301	11.6	382	1 PECH	pepsin A (EC 3.4.2)
9	299.5	11.6	396	2 A34401	cathepsin E (EC 3)
10	298.5	11.5	383	2 A41443	pepsin (EC 3.4.23)
11	296.5	11.5	412	1 KKHUO	cathepsin D (EC 3)
12	296	11.4	387	2 C38302	pepsin (EC 3.4.23)
13	296	11.4	391	2 A43356	cathepsin E (EC 3)
14	295	11.4	387	2 D38302	pepsin (EC 3.4.23)
15	295	11.4	407	1 KHRD	cathepsin D (EC 3)
16	291.5	11.3	388	1 S19684	pepsin A (EC 3.4.2)
17	287.5	11.1	444	2 T24204	hypothetical prote
18	286	11.1	387	2 E38302	pepsin (EC 3.4.23)
19	285.5	11.0	388	1 PEHU	pepsin A (EC 3.4.2)
20	285.5	11.0	388	1 PEMQAR	pepsin A (EC 3.4.2)
21	285.5	11.0	388	2 A30142	pepsin A (EC 3.4.2)
22	285.5	11.0	398	2 S66465	cathepsin E (EC 3)
23	285	11.0	389	2 JE0371	pepsin C (EC 3.4.2)
24	285	11.0	398	2 I51185	cathepsin D (EC 3)
25	284.5	11.0	410	1 KHMSD	cathepsin D (EC 3)
26	283.5	11.0	380	2 I47176	chymosin (EC 3.4.2)
27	283.5	11.0	388	1 PEMQAJ	pepsin A (EC 3.4.2)
28	282.5	10.9	388	2 B30142	pepsin A (EC 3.4.2)
29	281.5	10.9	381	1 CMSH8	chymosin (EC 3.4.2)

ALIGNMENTS

RESULT 1

A59090
aspartic proteinase (EC 3.4.23.-) BACE precursor - human
N:Alternate names: beta-secretase; beta-site APP cleaving enzyme
C:Species: Homo sapiens (man)
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 11-May-2000
C:Accession: A59090
R:Vassar, R.; Bennett, B.D.; Babu-Khan, S.; Kahn, S.; Mendiaz, E.A.; Denis, P.; Teplö
M.A.; Biere, A.L.; Curran, E.; Burgess, T.; Louis, J.C.; Collins, F.; Treanor, J.; Ro
Science 286, 735-741, 1999
A:Title: beta-Secretase cleavage of Alzheimer's amyloid precursor protein by the tran
A:Reference number: A59090; MUID:20002972
A:Note: submitted to GenBank, September 1999
A:Accession: A59090
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-501 <VAS>
A:Cross-references: GB:AF190725; NID:g6118538; PIDN:AAF04142.1; PID:g6118539
C:Genetics:
A:Gene: BACE
C:Superfamily: beta-secretase
C:Keywords: Alzheimer's disease; aspartic proteinase; brain; glycoprotein; hydrolase;
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-45/Domain: propeptide #status predicted <PRO>
F:46-501/Product: acid proteinase BACE #status predicted <MAT>
F:461-477/Domain: transmembrane #status predicted <TRN>
F:93-289/Active site: Asp #status predicted
F:153,172,223,354/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:330-380/Disulfide bonds: #status predicted

Query Match 100.0%; Score 2587; DB 2; Length 501;
Best Local Similarity 100.0%; Pred. No. 2.7e-201;
Matches 488; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGVLPAGTQHGIRLPLRSLGAGPLGLRLPRETDEEPEEGRGSGFVEMVNLKRSQ 60

Db 14 AGVLPAGTQHGIRLPLRSLGAGPLGLRLPRETDEEPEEGRGSGFVEMVNLKRSQ 73

QY 61 GYVVENTGSPQTLNLLVDTSNFAVCAAPHLHRYQRLSSTYRDLRKGVVVPYT 120

Db 74 GYVVENTGSPQTLNLLVDTSNFAVCAAPHLHRYQRLSSTYRDLRKGVVVPYT 133

QY 121 QKWEGELGTDLVSIPIPHGNVTVRANIAAITESDKFFINGSNWEGILGLAYAEIARPDOS 180

Db 134 QKWEGELGTDLVSIPIPHGNVTVRANIAAITESDKFFINGSNWEGILGLAYAEIARPDOS 193

QY 181 LEPPFDLSLVKQTHVPNLFSLQLCGAGFPLNQSEVLASVGSGMIIGDHSLTGSLWYTP 240

Db 194 LEPPFDLSLVKQTHVPNLFSLQLCGAGFPLNQSEVLASVGSGMIIGDHSLTGSLWYTP 253

QY 241 IRREWYEVIIIVRVEINGDOLKMDCKEYNDKSVGTTNLRPLPKKVFEEAAVKSIAAS 300

pepsin A (EC 3.4.2)
pepsinogen A - com
cathepsin E (EC 3)
chymosin (EC 3.4.2)
gastricsin (EC 3.4)
pepsin (EC 3.4.23)
aspartic proteinase
cathepsin D (EC 3)
prochymosin - comm
candidapepsin (EC
saccharopepsin (EC
aspartic proteinase
pepsinogen C - com
gastricsin (EC 3.4)
aspartic proteinase
gastricsin (EC 3.4)

```
|||||
Db 254 IREWEYEVIIIVRVEINGDQLKMDCKEYNDKSIDVSTTNLRPLPKKVFEEAVSIRKAAS 313
QY 301 STEKEPDGFWLGEQLVCWQAGTTPWNIIPVLSILMGVETNQSFRTILPQOYLRPVEDV 360
Db 314 STEKEPDGFWLGEQLVCWQAGTTPWNIIPVLSILMGVETNQSFRTILPQOYLRPVEDV 373
QY 361 ATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARRIGFAVSACHVHDEFFRTAAVEG 420
Db 374 ATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARRIGFAVSACHVHDEFFRTAAVEG 433
QY 421 PVTILMEDCGYNIPQTDSTILMTIAYMAAICALFMLPLCLMVCOMRCLRLRQHQHDDF 480
Db 434 PVTILMEDCGYNIPQTDSTILMTIAYMAAICALFMLPLCLMVCOMRCLRLRQHQHDDF 493
QY 481 ADDISLLK 488
Db 494 ADDISLLK 501

RESULT 2
JC7574
pepsinogen A - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 31-Mar-2001 #sequence_revision 31-Mar-2001 #text_change 31-Mar-2001
C:Accession: JC7574; PC7119
R:Ikuzawa, M.; Inokuchi, T.; Kobayashi, K.; Yasumasu, S.
J. Biochem. 129, 147-153, 2001
A:Title: Amphibian pepsinogens: Purification and characterization of Xenopus pepsinogens
A:Reference number: JC7573; PMID:11134969
A:Contents: Stomach
A:Accession: JC7574
A:Molecule type: mRNA
A:Residues: 1-384 <IKU>
A:Cross-references: DDBJ:AB045380
A:Accession: PC7119
A:Molecule type: protein
A:Residues: 16-35; 57-76 <IK2>
C:Comment: This protein is a zymogen for gastric aspartic proteinase, with pepsin-like a
C:Genetics:
A:Gene: Pga
C:Keywords: stomach; zymogen

Query Match 12.1%; Score 314; DB 3; Length 384;
Best Local Similarity 25.1%; Pred. No. 1.2e-17;
Matches 107; Conservative 69; Mismatches 154; Indels 96; Gaps 18;

QY 13 IRLPLRSLGGLGAPLGLRLPRETDEEPEEPGRGSRFVE-----MVDNLR 55
Db 17 VKVPLRKG-----ESFNRPRQLGLLDYLLKKNYPNASKYFPTLAQSSAETIQ 65
QY 56 KSGQGYVEMTVGSPPTLNLIVDTGSSNFNFAVGAHPFL-----HRYQROLSSPY 108
Db 66 NYMDIEYGTIGTTPQEFVIFDGSANLW---PSVYCSQACSNNRNPQSSSTF 122
QY 109 RDLRKGVVYPYTOGKWEGLGTLVSIHPGPNVTVRANIAATESDK--FFINGSNWEGIL 167
Db 123 QATNPVSIQYGTGSGMGLGYDTLQV---GNIQISNQMFGLSESEPGSFLYSPFDGIL 179
QY 168 GLAYAEIARDDSLFPFSLVQKTHVP--NLFSQLQCGAGFPLNQSEVLASVGGSMIIG 226
Db 180 GLAFPSIA--SSQATPVFNQMSQGLIPQNLFSVYLSDDG-----QTGSSVFLFG 227
QY 227 IDLSLTGSLWYTPIRREWEYEVIIIVRVEINGDQL--KMDCKEYNDKSIDVSTTNLR 284
Db 228 VDSNYSGSLNWVPLTAETWQITLDSVINGQVIACSSQSC-----QAIVDTGTSMTG 281
QY 285 PKVFEAAVKSIAASSTKFPDGLVGEOLV--CWQAGTTPWNIIPVLSILMGVETNQ 343
Db 282 PSTPI-ANIQNYIGASODSN-----GOVYINCNNISNMPTIVE----- 318
QY 344 FRITILPQOY-LRPVEDVATSDQDCYK-FAISQSSGTG---VMGAVIMEGFYVVFDRAR 397
```

```
|||||
Db 319 ---TINGVOYPLSPSAYVRQNOQCGSSGFOAMNLTNSGDLWILGDVFIROYFTVFDAN 375
QY 398 KRIGFA 403
Db 376 NYVAIA 381

RESULT 3
JC7575
pepsinogen A - bullfrog
C:Species: Rana catesbeiana (bullfrog)
C:Date: 31-Mar-2001 #sequence_revision 31-Mar-2001 #text_change 31-Mar-2001
C:Accession: JC7575
R:Ikuzawa, M.; Inokuchi, T.; Kobayashi, K.; Yasumasu, S.
J. Biochem. 129, 147-153, 2001
A:Title: Amphibian pepsinogens: Purification and characterization of Xenopus pepsinog
A:Reference number: JC7573; PMID:11134969
A:Contents: Stomach
A:Accession: JC7575
A:Molecule type: mRNA
A:Residues: 1-385 <IKU>
A:Cross-references: DDBJ:AB045376
C:Comment: This protein is a zymogen for gastric aspartic proteinase, with pepsin-lik
C:Genetics:
A:Gene: Pga
C:Keywords: stomach; zymogen

Query Match 11.9%; Score 308; DB 3; Length 385;
Best Local Similarity 27.8%; Pred. No. 3.7e-17;
Matches 99; Conservative 64; Mismatches 133; Indels 60; Gaps 15;

QY 62 YVEMTVGSPPTLNLIVDTGSSNFVAVG---AAPHPFLHRYQROLSSSTYRDLRKGYV 117
Db 73 YFGTISIGTTPQSFVIFDTGSSNLWVPSVYCSSPACTNNHNMENPQSSSTFOATNPVSI 132
QY 118 PYTOCKWEGELGTLVSIHPGPNVTVRANIAATESDK--FFINGSNWEGILGLAYAEIAR 176
Db 133 QYGTGMSGFLGYDTVQV---GNIQITNQIFGLSQSEPGSELYSPFDGLVGLAFPSLA- 188
QY 177 PDDSLPEPFDLSLVKQTHVP--NLFSQLQCGAGFPLNQSEVLASVGGSMIIGIDHSLSYTG 235
Db 189 -SSQATPVFNQMSQGLIPQNLFSVYLSQSG---QS-----GSFVLEGGVDTSYTG 237
QY 236 LWYTPIRREWEYEVIIIVRVEINGDQLK--DCKEYNDKSIDVSTTNLRPLPKVFEAAV 293
Db 238 LNWVPLTAETWQITVDSISIGGQVIACSGSC-----SAIVDTGTSLLAGP---STPI 287
QY 294 KSIRKAASSTKFPDGLVGEOLVQWQAGTTPWNIIPVLSILMGVETNQSFRTILPQOY 353
Db 288 ANIQYIYGANDSNGQYV---INCNNISNMPTVVF-----TINGVOY 326
QY 354 LRPVED-VATSDQDC---YKFAISQSSGTG---VMGAVIMEGFYVVFDRARRIGFA 403
Db 327 PLPASAYVRQSQSCTSGFOAMNLTSSGDLWILGDVFIROYFTVFDAN 382

RESULT 4
B38302
pepsin (EC 3.4.23.-) II-1 precursor - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 14-Jun-1991 #sequence_revision 20-Sep-1991 #text_change 23-Feb-1997
C:Accession: B38302
R:Kageyama, T.; Tanabe, K.; Koiwai, O.
J. Biol. Chem. 265, 17031-17038, 1990
A:Title: Structure and development of rabbit pepsinogens. Stage-specific zymogens, nu
A:Reference number: A38302; MUID:91009127
A:Accession: B38302
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-387 <KAG>
A:Cross-references: GB:M59235; GB:J05638
```

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

Db 159 TNOIFGLSETEPGFFLYFAPFDGILGLAYPSIS--SSGATPVFDNIWNORLVSODLFSVY 216

Db 75 TPPQNFLVLFDTGSSNLWT

```

75  T P P O N F L V L E D T G S S N L W A S T Y C O S O A C T N H P L - - - - F N P S O S S T Y S S N O O O F S L

```

75 T P P O N F L V L F D T G S S N L W V A S T Y C O S O A C T N H P L - - - - F N P S O S S T Y S S N O O O F S L O Y G T 130

QY 202 LCAGFPLNQSEVLASVGGSMIGGIDHSLYTGSLWYTPIRREWYVEVLIIVVEINGDGL 261
DB 217 LSAD-----DQS-----GSWIFGGIDSSYTGSLNWVPSVEGWQISVDSITMNGKTI 266
QY 262 --KMDCKEYNDKSIYDSTGTLRLPKKVFEEAAVKSIAKASTERKFPDGFWMGEQLV-CW 318
DB 267 ACAKGC-----QAIVDTGTSLLTGTPSIANTQSDIGASENSD-----GENVSCS 312
QY 319 QAGTTPWNIFFVVISLYLMGEVNTNQSPRIITLPQY-LRPVEDVATSQDDCYK-----FAI 372
DB 313 AISSLPDIVF-----TINGVQYPLPPSAVILQSGSCTSGFQGMQDVP 354
QY 373 QSSTGTVMGAVIMEGFYVVFDRARRIGFA 403
DB 355 TESGELMILGDVFIQYFTVDRANNOVGLA 385

RESULT 7
A39314
gastricsln (EC 3.4.23.3) precursor - bullfrog
C:Species: Rana catesbeiana (bullfrog)
C:Date: 19-Jun-1992 #sequence_revision 19-Jun-1992 #text_change 22-Jun-1999
C:Accession: A39314
R:Yakabe, E.; Tanji, M.; Ichinose, M.; Goto, S.; Miki, K.; Kurokawa, K.; Ito, H.; Kageya
J. Biol. Chem. 266, 22436-22443, 1991
A:Title: Purification, characterization, and amino acid sequences of pepsinogens and pep
A:Reference number: A39314; MUID:92042186
A:Accession: A39314
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-384 <YAK>
A:Cross-references: GB:M73750; NID:g213687; PIDN:AAA49530.1; PID:g213688
C:Superfamily: pepsin
C:Keywords: aspartic proteinase; hydrolase; protein digestion

Query Match 11.7%; Score 302; DB 2; Length 384;
Best Local Similarity 24.5%; Pred. No. 1.1e-16;
Matches 105; Conservative 65; Mismatches 147; Indels 112; Gaps 17;

QY 11 HGIRLPLRSLGGAPLGLRLPRETDEPEEPGR--GSFVEMVDNLRGKSGQGYVEMTV 68
DB 35 HGIRAPV-----VDPATKYNNFATAFEPLANVMDMSYGEISI 73
QY 69 GSPPTLNLIVDTGSSNFVGAAPHPL-----HRYQRLSSYTRDLRKGVVYPTQ 121
DB 74 GTPQNFVLFDTGSSNLW---PSTYCQOACTNHPQFNPSQSSVSSNQOQFSLQYGT 130
QY 122 KWGEGLTDLVSPHGNVTVRANIA-----AITESDKFFINGSNWEGILGLAYAE 173
DB 131 GSLTGILGYDTVQI-----QNAISQOEFGLSVTEPTNFVY-AQPDGILGLAYS 180
QY 174 IARPDLSLEPFDFSLVKQTHVPN-LFSLQCGAGFPLNQSEVLASVGGSMIIGIDHSLY 232
DB 181 IA--EGGATTVQGMQIQLNQLPFAFYLGGQNSQ-----GGEVAFGVDQNY 230
QY 233 TGSWYTPIRREWYVEVLIIVVEINGD-----LMDCKEYNDKSIYDSTGTLRLPKKVF 289
DB 231 SGQIYWTPTSETYQWIGIOGFSVNGQATGWCQSGC-----QGIVDTGTSLLTAPQSVF 284
QY 250 EAAVKSIAKASTEKFPDGFWMGEQLV-CWQAGTTPWNIFFV-----SLYLMGEV 340
DB 285 SSLMOSIGAQQDN-----GQAVSCSNIOQLPISITISGVSPPLPPSAVILQONS 336
QY 341 NO---SFRITILPQOYLRPVEDVATSDDCYKFAISOSSSTGTVMGAVIMEGFYVVFDRAR 397
DB 337 GYCITIGIMPTYLPSONGQPL-----WILGDVFLRQYYSVYDLGN 375
QY 398 KRIGFAVA 406
DB 376 NQVGFAAAA 384

RESULT 8
PECH
Pepsin A (EC 3.4.23.1) precursor - chicken
N:Alternate names: pepsinogen A
C:Species: Gallus gallus (chicken)
C:Date: 18-Apr-1984 #sequence_revision 01-Dec-2000 #text_change 01-Dec-2000
C:Accession: JE0370; A00984
R:Sakamoto, N.; Saiga, H.; Yasugi, S.
Biochem. Biophys. Res. Commun. 250, 420-424, 1998
A:Title: Analysis of temporal expression pattern and cis-regulatory sequences of chic
A:Reference number: JE0370; MUID:98440813
A:Accession: JE0370
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-382 <SAK>
A:Cross-references: GB:AB025281; NID:g4589837; PIDN:BAA76891.1; PID:g4589838
R:Baudys, M.; Kostka, V.
Eur. J. Biochem. 136, 89-99, 1983
A:Title: Covalent structure of chicken pepsinogen.
A:Reference number: A00984; MUID:84004412
A:Accession: A00984
A:Molecule type: protein
A:Residues: 16-87, S, 89-382 <BAU>
C:Superfamily: pepsin
C:Keywords: aspartic proteinase; gastric juice; glycoprotein; hydrolase; protein dige
F:16-57/Domain: activation peptide #status experimental <APT>
F:92-382/Product: pepsin A #status predicted <MAT>
F:92-275/Active site: Asp #status predicted
F:105-110, 266-270, 305-338/Disulfide bonds: #status experimental
F:128/Binding site: carbonylate (Asn) (covalent) #status experimental

Query Match 11.6%; Score 301; DB 1; Length 382;
Best Local Similarity 24.0%; Pred. No. 1.4e-16;
Matches 88; Conservative 68; Mismatches 126; Indels 84; Gaps 13;

QY 62 YVEMTVGSPPTLNLIVDTGSSNFVGAAPHPL-----HRYQRLSSYTRDLRK 114
DB 74 YGTISIGTPQODFTVIFDTGSSNLW---PSYCKSSACSNHKKRDFSKSYVSTNET 130
QY 115 VVVPYTGKWEGLTDLVSPHGNVTVRANIAAITESDK--FFINGSNWEGILGLAYAE 173
DB 131 VVIAGTGSMSGILGYDTVAV---SSIDVQNIQFGLSETEPGSFYFCNFDFILGLAFPS 187
QY 174 IARPDLSLEPFDFSLVKQTHVPN-LFSLQCGAGFPLNQSEVLASVGGSMIIGIDHSLY 232
DB 188 IS--SSGATPVFDNMMSQHLVAQDLFSVYLSKDG-----ETGSFVLFGGIDPNYT 235
QY 233 TGSWYTPIRREWYVEVLIIVVEINGDGL--MDCKEYNDKSIYDSTGTLRLPKKVF 290
DB 236 TKGIVWPVLSAETVQWITMDRVTVGNKYVACFFTC-----QAIVDTGTSLLVMPQGA 289
QY 291 AAVKSIKAASSTE-----KFPDGFWMGEQLVQWQAGTTPWNIFFVVISLYLMGEVNTQS 343
DB 290 RIKDLGVSDGEISDIDSKLPD-----VTFFINGHA---- 322
QY 344 FRITILPQOYLRPVEDVATSDDCYKFAISOSSSTGT-----VMGAVIMEGFYVVFDRAR 397
DB 323 -----FTLPASAVLNEDGSCMLGFENMGTPTELGEQWILGDVFLREYVIFDRAN 373
QY 398 KRIGFA 403
DB 374 NKVGLS 379

RESULT 9
A34401
cathepsin E (EC 3.4.23.34) precursor - human
C:Species: Homo sapiens (man)
C:Date: 22-Jun-1990 #sequence_revision 22-Jun-1990 #text_change 22-Jun-1999
C:Accession: A42038; A34401; S35663; S34467; A34643; B34643
R:Azuma, T.; Liu, W.; Vander Laan, D.J.; Bowcock, A.M.; Taggart, R.T.
J. Biol. Chem. 267, 1609-1614, 1992

A::Title: Human gastric cathepsin E gene. Multiple transcripts result from alternative p
A::Reference number: A42038; MUID:92112877
A::Accession: A42038
A::Molecule type: DNA
A::Residues: 1-396 <AZ>
A::Cross-references: GB:M84424; GB:M82847; NID:g181203; PIDN:AAA52300.1; PID:g181205
R:Aote: sequence extracted from NCBI backbone (NCBIN:75963, NCBIN:75966, NCBIN:75971,
R:Asuma, T.; Pals, G.; Mohandas, T.K.: Couvreux, J.M.; Taggart, R.T.
J. Biol. Chem. 264, 16748-16753, 1989
A::Title: Human gastric cathepsin E. Predicted sequence localization to chromosome 1, and
A::Reference number: A34401; MUID:89380302
A::Accession: A34401
A::Molecule type: mRNA
A::Residues: 1-396 <AGZ>
A::Cross-references: GB:J05036; NID:g181193; PIDN:AAA52130.1; PID:g181194
R:Takeda-Ezaki, M.; Yamamoto, K.
Arch. Biochem. Biophys. 304, 352-358, 1993
A::Title: Isolation and biochemical characterization of procathepsin E from human erythrocytes.
A::Reference number: S35663; MUID:93349047
A::Accession: S35663
A::Status: preliminary
A::Molecule type: protein
A::Residues: 20-38;54-76 <TAK>
F:Hill, J.; Montgomery, D.S.; Kay, J.
FEBS Lett. 326, 101-104, 1993
A::Title: Human cathepsin E produced in E. coli.
A::Reference number: S34467; MUID:93314762
A::Accession: S34467
A::Status: preliminary
A::Molecule type: protein
A::Residues: 57-60,62-81 <HIL>
R:Achanda, S.B.P.; Matsuzaki, O.; Kageyama, T.; Takahashi, K.
Biochem. Biophys. Res. Commun. 168, 878-885, 1990
A::Title: Structural evidence for two isozytic forms and the carbohydrate attachment site
A::Reference number: A34643; MUID:90241267
A::Accession: A34643
A::Status: preliminary
A::Molecule type: protein
A::Residues: 54-58,'XXX','M',66-89,'X','91-95 <ATH>
A::Accession: B34643
A::Status: preliminary
A::Molecule type: protein
A::Residues: 54-59,'X',61-68 <ATZ>
C:Genetics:
A::Gene: GDB:CTSE
A::Cross-references: GDB:I19821; OMIM:116890
A::Map position: lq31-lq31
C:Superfamily: peptin
C:Keywords: aspartic proteinase; blocked amino end; hydrolase; zymogen
F:1-17/Domin: signal sequence #status predicted <SIG>
F:18-53/Domin: activation peptide #status predicted <PRO>
F:54-396/Product: cathepsin E #status predicted <MAT>
F:18/Modified site: blocked amino end (Gln) (in mature form) (probably pyrrolidone carbonyl)
F:96,281/Active site: Asp #status predicted

Query Match 11.6%; Score 299.5; DB 2; Length 396;
Best Local Similarity 25.9%; Pred No. 1.9e-16;
Matches 100; Conservative 65; Mismatches 148; Indels 73; Gaps 16;

Oy 35 DEEPEEPGRSGVEMVDNLKRGSCQGYVEVMYVGSPPTLNILVDTGSSNFAYGA---- 90
 :
Db 63 QDSAKEP-----LNYLD-----MEVFGTISIGSPPQNFTVIDTGSSNLLWVPVSVCVT 110

Oy 91 APHFELHYQRQLSYRDLRKGYYPVITCKKEGLDGLTDLSVIPHPNVTVRANTAAI 150
 : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 111 SPACKTHSRFPQSSTYSQPQGFSIQYTGGSLGIADQSV--EGLTIVGVQQFGESV 169

Oy 151 TESDKFFNGSNREGIIGLVAETARPDSDLEPFOSLVKQTHVPNLFSLQCCAGGPLN 210
 || : : : : : : : : : : : | : : | : : : : | : : | : : | : :| : :
Db 170 TEPGQT FVD-AEFDGILLGLGPSLA---VGGVTPVPFDNNMAQ-----NLVDLPMFVSVMSSN 222

Oy 211 QSEVLASVGGSMIIGGDHSLYTSGLWTFPIRRWEVYEVIIVRVINGODLKMDCKEYN 270

A:Title: Cloning and sequence analysis of cDNA for human cathepsin D.
A:Reference number: A25771; MUID:85270436
A:Accession: A25771
A:Molecule type: mRNA
A:Residues: 1-412 <FAU>
A:Cross-references: EMBL:M11233; NID:g181179; PIDN:AAB59529.1; PID:g181180
R:Westley, B.R.; May, F.E.B.
Nucleic Acids Res. 15, 3773-3786, 1987
A:Title: Oestrogen regulates cathepsin D mRNA levels in oestrogen responsive human breast
A:Reference number: S30749; MUID:87231068
A:Accession: S30749
A:Molecule type: mRNA
A:Residues: 1-412 <WES>
A:Cross-references: EMBL:X05344; NID:g29677; PIDN:CAA28955.1; PID:g29678
R:May, F.E.B.; Smith, D.J.; Westley, B.R.
Gene 134, 277-282, 1993
A:Title: The human cathepsin D-encoding gene is transcribed from an estrogen-regulated enhancer
A:Reference number: PC2066; MUID:94085791
A:Accession: PC2066
A:Molecule type: DNA
A:Residues: 1-23 <MAY>
A:Cross-references: GB:L12980; NID:g291930; PIDN:AAAL6314.1; PID:g455429
A:Experimental source: MCF-7 cell
R:Cavallies, V.; Augereau, P.; Rochefort, H.
Proc. Natl. Acad. Sci. U.S.A. 90, 203-207, 1993
A:Title: Cathepsin D gene is controlled by a mixed promoter, and estrogens stimulate only one promoter
A:Reference number: I59236; MUID:93126342
A:Accession: I59236
A:Status: translation not shown; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-22 <CAV1>
A:Cross-references: GB:I52557; NID:g263124; PIDN:AAD13868.1; PID:g4261568
R:Augereau, P.; Miralles, F.; Cavallies, V.; Gaudelet, C.; Parker, M.; Rochefort, H.
Mol. Endocrinol. 8, 693-703, 1994
A:Title: Characterization of the proximal estrogen-responsive element of human cathepsin D
A:Reference number: I57716; MUID:95021301
A:Accession: I57716
A:Status: translation not shown; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-22 <CAV2>
A:Cross-references: GB:S74689; NID:g786350; PIDN:AAD14156.1; PID:g4261856
R:Baldwin, E.T.; Bhat, T.N.; Gulnik, S.; Erickson, J.W.
submitted to the Brookhaven Protein Data Bank, April 1993
A:Reference number: A51839; PDB:1LYA
A:Contents: annotation; X-ray crystallography, 2.5 angstroms, residues 65-161; 170-241
R:Baldwin, E.T.; Bhat, T.N.; Gulnik, S.; Erickson, J.W.
submitted to the Brookhaven Protein Data Bank, April 1993
A:Reference number: A51840; PDB:1LYB
A:Contents: annotation; X-ray crystallography, 2.5 angstroms, with inhibitor residues 65-161; 170-241
R:Baldwin, E.T.; Bhat, T.N.; Gulnik, S.; Rosur, M.V.; Sowder II, R.C.; Cachau, R.E.; Collier, P.
Proc. Natl. Acad. Sci. U.S.A. 90, 6796-6800, 1993
A:Title: Crystal structures of native and inhibited forms of human cathepsin D: Implications for the catalytic mechanism
A:Reference number: A48229; MUID:93342076
A:Contents: annotation; X-ray crystallography, 2.5 angstroms
C:Comment: Cathepsin D is a ubiquitous lysosomal proteinase.
C:Comment: In addition to the propeptide, residues 163-168 and 411-412 are proteolytically cleaved.
C:Comment: The carbohydrate bound to 134-Asn contains a mannose-6-phosphate that is bound to the protein.
C:Genetics:
A:Gene: GDB:CTSD
A:Cross-references: GDB:I20512; OMIM:116840
A:Map position: 11p15.5-11p15.5
C:Function:
A:Description: limited specificity endopeptidase
A:Pathway: intracellular protein degradation
C:Superfamily: peptin
C:Keywords: aspartic proteinase; glycoprotein; hydrolase; lysosome; protein degradation
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-64/Domain: propeptide #status predicted <PRO>
F:65-162/Region: signal sequence #status predicted <SIG>
F:163-236/Region: phosphatransferase recognition
F:237-329/Region: phosphatransferase recognition
F:330-412/Region: phosphatransferase recognition
F:91-160/110-117,286-290,329-366/Disulfide bonds: #status experimental
F:97,295/Active site: Asp #status experimental
F:134,263/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 11.5%; Score 296.5; DB 1; Length 412;
Best Local Similarity 26.8%; Pred. No. 3.5e-16;
Matches 117; Conservative 66; Mismatches 163; Indels 91; Gaps 19;
QY 13 IRLPLR-----SGLGGAPLGL-----RLPRETDEEPEEGRGRGSFVEMVD 52
DB 22 VRIPLHKFTSIRRTMSEVGGSDIEDIAKGPVSKYQAVPAVTE-----GPIPEVLK 72
QY 53 NLRGKGGGYVYEMTVGSPPTNLIVDTGSSNFVAGAAHPFL-----HRYRQRQSS 106
DB 73 NYMDAQ---YGEIGIGTPPQCFTVVFDTGSSNLWVPSIHCLLDTACWIHHKYNDSKSS 129
QY 107 TYRDLRKGVVYPYTOGKWEGLGTDLVSTP-----HGNVTVVRANIAAITESDKFF 157
DB 130 TVVKNGTSPDIHYGGSLGYSQDTVSVCOSASASALGQVKVERQVFGATKQPGIT 189
QY 158 INGSNMEGILGLAYAEIARPDSDLEFFDLSLVKQTHV--PNLFSLOLCGAGFPLNOSVLA 216
DB 190 FTAARFDGILGMAYPRIS--VNNVLPVFDNLMOQLVDQNIQFVYL-----SRDPDA 239
QY 217 SVGGSMIIGGIDHSLSYTGSLWYTPTRREWHYEVIIIVVEI--NGQDLKMDCKEYNDKSI 275
DB 240 QPGGELMLGGTDSKYKGLSLSYLVNTRKAYQVHLDQVEVASGLTL---CKE--GCEAIV 294
QY 276 DSGTTLNLRLPKKVFEEAAVKSIKAASSTEFKPDGFWLGEOLV--CWQAGTTPWNIFPVISLY 334
DB 295 DNGTSLMVGPDVEVRELQKAGVPLIQ-----GEWIPCEKVST-----LPAITLK 341
QY 335 LMGEVYNSFRITILPQOYLRPVEDVATSDQDCYKFAISQ-----SSTGTVMGAVIMEGF 389
DB 342 LGG---KGYKLS--PEDYTLKVSQAGKTL--CLSGFMGMDIPPPSGPLWILGDFVIGRY 393
QY 390 YVVDPRKRKRIGFAYSA 406
DB 394 YTVFDRNNRVGFEEAA 410
RESULT 12
C38302
pepsin (EC 3.4.23.-) II-2/3 precursor - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 14-Jun-1991 #sequence_revision 14-Jun-1991 #text_change 23-Feb-1997
C:Accession: C38302
R:Kageyama, T.; Tanabe, K.; Koiwai, O.
J. Biol. Chem. 265, 17031-17038, 1990
A:Title: Structure and development of rabbit pepsinogens. Stage-specific zymogens, nucleotide sequence, and properties of the mature enzyme
A:Reference number: A38302; MUID:91009127
A:Accession: C38302
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-387 <KAG>
A:Cross-references: GB:J05638
C:Superfamily: peptin
C:Keywords: aspartic proteinase; hydrolase; phosphoprotein; protein digestion
Query Match 11.4%; Score 296; DB 2; Length 387;
Best Local Similarity 26.9%; Pred. No. 3.5e-16;
Matches 97; Conservative 63; Mismatches 135; Indels 66; Gaps 13;
QY 62 YVEMTVGSPPTNLIVDTGSSNFVAGAAHPF-----LHRYRQRQSSYRDLRKG 114
DB 75 YFGTISIGTPPDFTVIFDTGSSNLWV---PSTYCGSSLACALHKKRPNEDSSTYQCTSET 131
QY 115 VVVPYTOGKWEGLGTDLVSIIPHGNVTVVRANIAAITESDKFFINGSNWEGTLGLAYAEI 174
DB 132 LSITVTGSMTGILGYDTPVKVGSIEDTNOIFGLSKTEPSLTLF--APFDGLGLAYPSI 189
QY 175 ARPDSLPEFFDLSLVKQTHV--PNLFSLOLCGAGFPLNOSVLAISVGGSMIIGGIDHSLSY 233
DB 190 SSSDAT--PVFDNMNNEGLVSDLFVSVLSDD-----EKGSLVMFGGIDSSVYT 237

QY	234	GS	LWYTP	PIREWY	YEV	IIV	VEING	QDLK	M	-	DK	KEYN	YDK	SIV	DSCT	TN	LR	LP	PKK	V	F	E	A	291
			:	:	:	:		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
Db	238	GS	LMV	VP	SV	YEG	WQIT	MD	S	V	S	I	N	G	E	T	I	A	C	A	D	S	C	287
			:	:	:	:		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
		---	Q	A	I	V	D	T	C	T	S	L	L	T	G	P	---	---	---	---	---	---	---	---
QY	292	AV	K	S	I	K	A	A	S	T	E	R	F	P	D	G	F	W	L	G	E	Q	L	350
			:	:	:	:		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
Db	288	A	I	S	I	O	S	I	G	A	S	K	---	---	N	L	G	E	N	V	I	S	C	325
			:	:	:	:		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
		---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
QY	351	Q	O	L	R	P	V	E	D	V	A	T	S	O	D	C	K	F	A	I	S	O	S	404
			:	:	:	:		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
Db	326	I	O	P	L	P	A	S	A	I	L	K	E	D	D	C	T	S	G	L	E	C	M	385
			:	:	:	:		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
		---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
QY	405	S	405																					
Db	386	A	386																					

RESULT 13

A43356

cathepsin E (EC 3.4.23.34) precursor - guinea pig

N;Alternate names: erythrocyte membrane aspartic proteinase; slow-moving proteinase

C;Species: Cavia porcellus (guinea pig)

C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 22-Jun-1999

C;Accession: A43356

R;Kageyama, T.; Ichinose, M.; Tsukada, S.; Miki, K.; Kurokawa, K.; Koiwai, O.; Tanji, M.

J. Biol. Chem. 267, 16450-16459, 1992

A;Title: Gastric procathepsin E and progastricsin from guinea pig. Purification, molecular cloning, and cDNA sequence determination.

A;Reference number: A43356; MUID:92355614

A;Accession: A43356

A;Molecule type: mRNA

A;Residues: 1-391 <RAG>

A;Cross-references: GB:M88653; NID:g191294; PIDN:AAA37052.1; PID:g191295

A;Note: sequence extracted from NCBI backbone (NCBI:110763, NCBI:110769)

C;Superfamily: pepsin

C;Keywords: aspartic proteinase; hydrolase; membrane protein

Query Match

11.4%; Score 296; DB 2; Length 391;

Best Local Similarity 26.9%; Pred. No. 3.6e-16;

Matches 98; Conservative 64; Mismatches 130; Indels 72; Gaps 16;

QY

62

YV

E

W

T

V

G

S

P

P

Q

T

L

N

I

L

V

D

T

G

S

S

N

F

A

V

G

A

A

P

H

P

L

H

R

Y

Q

R

L

S

S

T

Y

R

D

L

R

K

G

V

Y

117

Db

74

Y

F

G

T

I

S

G

P

P

Q

N

F

T

V

I

D

T

G

S

S

N

L

V

P

S

V

C

T

P

S

P

A

C

O

T

H

P

V

F

H

P

S

L

S

T

Y

R

E

V

G

N

S

F

S

I

133

QY

118

P

T

O

G

K

E

G

E

L

G

T

D

L

S

I

P

H

G

P

N

T

V

R

A

N

T

A

A

T

E

S

D

K

F

F

I

N

G

S

N

W

E

I

L

G

L

A

Y

A

E

T

A

R

P

177

Db

134

Q

Y

G

T

S

L

T

I

G

A

D

O

V

S

E

S

L

T

V

G

O

F

G

E

S

O

E

P

K

T

F

V

A

E

F

D

A:Title: Isolation and sequencing of a cDNA clone encoding rat liver lysosomal cathepsin
A:Reference number: JQ1177; MUID:91354249
A:Accession: JQ1177
A:Molecule type: mRNA
A:Residues: 1-14, 'A', '16-204, 'N', '206-261, 'N', '263-407 <FUJ>
A:Accession: PQ0222
A:Molecule type: protein
A:Residues: 65-74;118-127;165-174 <FU2>
A:Experimental source: liver
A:Comment: Cathepsin D in rat liver lysosome occurs as a mixture of both a single chain
C:Function:
A:Description: limited specificity endopeptidase
A:Pathway: intracellular protein degradation
C:Superfamily: pepsin
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-64/Domain: propeptide #status predicted <PRO>
F:65-407/Product: cathepsin D, 43K single-chain form #status predicted <MAT>
F:65-164/Product: (or 65-165) cathepsin D 12K light chain #status predicted <MA2>
F:65-117/Product: cathepsin D 9K light chain #status predicted <MA4>
F:118-407/Product: cathepsin D 34K heavy chain #status predicted <MA5>
F:165-407/Product: (or 166-407) cathepsin D 30K heavy chain #status predicted <MA3>
F:91-160,110-117,281-285,324-361/Disulfide bonds: #status predicted
F:97,290/Active site: Asp #status predicted
F:134,258/Binding site: carbonylurate (Asn) (covalent) #status predicted

Query Match 11.4%; Score 295; DB 1; Length 407;
Best Local Similarity 26.4%; Pred. No. 4.6e-16;
Matches 114; Conservative 72; Mismatches 160; Indels 86; Gaps 19;

QY 13 IRLPLR-----SGLGA--PLGLRLPRETDEEPGRGRSGFVEMVNDLRKSKQG 61
DB 22 IRLPLRKTSTIRNMTVEGGSVEDLILKPTIKYSMQSPRTKEPVSELLKNLYDAQ--- 78
QY 62 YVEMTVGSPQTLNILDVGTSSNFAVGAAPHFL-----HRYQRLSSSTYRDLRKGV 115
DB 79 YGEIGIGTPPCFTVVFDTGSSNLWVPSIHCKLLDIACVWHHKYNSDKSSTYVKNGTSTF 138
QY 116 YVPTQGWKEGLGDLVSIPIHGNVTVRANTAAITESDKFF-----INGSNWEG 165
DB 139 DIHYGSGSLGVLSDQTVSVP-----CKSDLGKIKVEKQIFGEATKQGVVVFIAAKFDG 192
QY 166 ILGLAYAEIARPDSDLPEFFDSLVKQTHV-PNLFSIQLCGAGFPLNQSEVLASVGSMMII 224
DB 193 ILGMGYPFIS--VNKVLVPVDFNLMKQKLVKNIFS-----FYLNR-DPTGQPGGELML 242
QY 225 GGIDHSLYTGSLWYTPIRREWYEVIIIVRVEINGODLRMDCKEYNWYDKSIVDSGTTNLR 284
DB 243 GGTDSRYHGYELSYLVNVRKAYQVHMDQLEV-GSELT-CK--CGCAIVDTGTSLLVG 298
QY 285 PKKVFEEAAVKSIAASSTKEFPDGFGLGEQLV-CWQAGTTPWNIPFVLSLYLMGEVTNQ 343
DB 299 PVDEVKELQKAGVPLIQ-----GEYMIPECVKSS-----LPIITFKLGQ----- 340
QY 344 FRITILPOOYLRPVEDVATSDQDCYKFAIS-----QSTGTVMGAVIMEGYVYVFD 394
DB 341 -NYELHPEKYLKVSQAGKT-----ICLSGFMGMDIPPPSGPLWILGDFVIGCYITVFD 393
QY 395 RARKRIGFAVSA 406
DB 394 REYNRVGFAKAA 405

Search completed: September 6, 2001, 16:45:45
Job time: 489 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:37:36 ; Search time 72.75 Seconds
(without alignments)
229.783 Million cell updates/sec

Title: US-09-603-713-2

Perfect score: 2587

Sequence: 1 AGVLPAGHTGIRPLRSG.....CLRLRQHDDFADDISLLK 488

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	2587	100.0	501	1	BACE_HUMAN
2	2519	97.4	501	1	BACE_RAT
3	2484	96.0	501	1	BACE_MOUSE
4	1175	45.4	518	1	BAE2_HUMAN
5	327	12.6	324	1	PEP1_GADMO
6	306	11.8	387	1	PEP1_RABIT
7	304.5	11.8	388	1	PEP4_MACFU
8	302	11.7	367	1	PEP4_CHICK
9	299.5	11.6	396	1	CATE_HUMAN
10	298.5	11.5	383	1	PEPE_CHICK
11	296.5	11.5	412	1	CATD_HUMAN
12	296	11.4	387	1	PEP2_RABIT
13	296	11.4	391	1	CATE_CAVPO
14	295	11.4	387	1	PEP4_RABIT
15	295	11.4	407	1	CATD_RAT
16	291.5	11.3	388	1	PEP2_MACFU
17	286	11.1	387	1	PEP3_RABIT
18	285.5	11.0	388	1	PEP4_HUMAN
19	285.5	11.0	388	1	PEP4_MACMU
20	285.5	11.0	398	1	CATE_RAT
21	285	11.0	398	1	CATD_CHICK
22	284.5	11.0	410	1	CATD_MOUSE
23	283.5	11.0	388	1	PEP1_MACFU
24	281.5	10.9	381	1	CHYM_SHEEP
25	278.5	10.8	386	1	PEP4_PIG
26	277.5	10.7	396	1	CATE_RABIT
27	277.5	10.7	397	1	CATE_MOUSE
28	276.5	10.7	419	1	CARV_CANAL
29	273.5	10.6	381	1	CHYM_BOVIN
30	273	10.6	388	1	PEPE_RABIT
31	270.5	10.5	377	1	PEP2_MACFU
32	270	10.4	376	1	PAG2_BOVIN
33	266	10.3	405	1	CARP_YEAST

34	264	10.2	396	1	CARP_NEUCR	Q01294	neurospora
35	263	10.2	394	1	PEPC_CAVPO	Q64411	cavia porce
36	261.5	10.1	387	1	ASPP_AEDAE	Q03168	aedes aegypt
37	261.5	10.1	388	1	PEPC_HUMAN	P20142	homo sapien
38	261	10.1	388	1	PAG_HORSE	Q28389	equus caball
39	260	10.1	345	1	CATD_PIG	P00795	sus scrofa
40	258.5	10.0	496	1	ASPR_ORYSA	P42211	oryza sativ
41	253	9.8	392	1	PEPC_RAT	P04073	rattus norv
42	251.5	9.7	389	1	PAG1_PIG	Q29078	sus scrofa
43	250	9.7	402	1	RENI_RAT	P08424	rattus norv
44	250	9.7	406	1	RENI_HUMAN	P00797	homo sapien
45	249.5	9.6	509	1	APR1_ORYSA	Q42456	oryza sativ

ALIGNMENTS

```

RESULT 1
BACE_HUMAN
ID BACE_HUMAN STANDARD; PRT: 501 AA.
AC P56817; Q9UJTS;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE BETA-SECRETASE PRECURSOR (EC 3.4.23.-) (BETA-SITE APP CLEAVING ENZYME)
DE (BETA-SITE AMYLOID PRECURSOR PROTEIN CLEAVING ENZYME) (ASPARTYL
DE PROTEASE 2) (ASP 2) (ASP2) (MEMBRANE-ASSOCIATED ASPARTIC PROTEASE 2)
DE (MEMAPSN-2).
GN BACE OR BACE1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20002972; PubMed=10531052;
RA Vassar R., Bennett B.D., Babu-Khan S., Kahn S., Mendiaz E.A.,
RA Denis P., Teplow D.B., Ross S., Amarante P., Loeloff R., Luo Y.,
RA Fisher S., Fuller J., Edenson S., Lile J., Jarosinski M.A.,
RA Biere A.L., Curran E., Burgess T., Louis J.-C., Collins F.,
RA Treanor J., Rogers G., Citron M.;
RA "Beta-secretase cleavage of Alzheimer's amyloid precursor protein by
RT the transmembrane aspartic protease BACE.";
RL Science 286:733-741(1999).
RN [2]
RP SEQUENCE FROM N.A., SEQUENCE OF 46-68, AND CHARACTERIZATION.
RC TISSUE=Brain;
RX MEDLINE=20057171; PubMed=10591214;
RA Sinha S., Anderson J.P., Barbour R., Basl G.S., Caccavello R.,
RA Davis D., Doan M., Doney H.F., Frigon N., Hong J., Jacobson-Croak K.,
RA Jewett N., Keim P., Knops J., Lieberburg I., Power M., Tan H.,
RA Tatsuno G., Tung J., Schenk D., Seubert P., Suomensari S.M., Wang S.,
RA Walker D., Zhao J., McConlogue L., Varghese J.;
RA "Purification and cloning of amyloid precursor protein beta-secretase
RT from human brain.";
RL Nature 402:537-540(1999).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=20057170; PubMed=10591213;
RA Yan R., Bienkowski M.J., Shuck M.E., Miao H., Tory M.C., Pauley A.M.,
RA Brashier J.R., Stratman N.C., Mathews W.R., Buhl A.E., Carter D.B.,
RA Tomasselli A.G., Parodi L.A., Heinrichson R.L., Gurney M.E.;
RA "Membrane-anchored aspartyl protease with Alzheimer's disease beta-
RT secretase activity.";
RL Nature 402:533-537(1999).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=20030166; PubMed=10561122;
RA Hussain I., Powell D.J., Howlett D.R., Tew D.G., Meek T.D.,
RA Chapman C., Gloger I.S., Murphy K.E., Southan C.D., Ryan D.M.,
RA Smith T.S., Simons D.L., Walsh F.S., Dingwall C., Christie G.;
RA "Identification of a novel aspartic proteinase (Asp 2) as beta-
RT

```

Db	194	LEPFFDLSVKQTHVNPVLSQLCGAGPFLNDOSEVLASVGGSGSMITGGIDHSITGSLWTP	253																																																																																																																																																																																																																																																
QY	241	IRREWYEVIIIVRVEINGODLKMDCKEYNDKSIYDVGTTNLRPKKVFEEAAVKSIAAS	300																																																																																																																																																																																																																																																
Db	254	IRREWYEVIIIVRVEINGODLKMDCKEYNDKSIYDVGTTNLRPKKVFEEAAVKSIAAS	313																																																																																																																																																																																																																																																
QY	301	STEKFPDGFWMGEQLVCMQAGTTPWNIIPVLSILYLMGEVTTNQSFRITILLPOOYLRPVEDV	360																																																																																																																																																																																																																																																
Db	314	STEKFPDGFWMGEQLVCMQAGTTPWNIIPVLSILYLMGEVTTNQSFRITILLPOOYLRPVEDV	373																																																																																																																																																																																																																																																
QY	361	ATSQDDCYKFAISQSSTGTVMGAVTMEGFYVVDERRARRKRGFAVSACHVHDEFRTAAVEG	420																																																																																																																																																																																																																																																
Db	374	ATSQDDCYKFAISQSSTGTVMGAVTMEGFYVVDERRARRKRGFAVSACHVHDEFRTAAVEG	433																																																																																																																																																																																																																																																
QY	421	PFVTLDMEDCGYNIPQTDSTELMTIAYVMAATCALFMPLCLMVCQWCLRLCRLQHQHDF	480																																																																																																																																																																																																																																																
Db	434	PFVTLDMEDCGYNIPQTDSTELMTIAYVMAATCALFMPLCLMVCQWCLRLCRLQHQHDF	493																																																																																																																																																																																																																																																
QY	481	ADDISLLK 488																																																																																																																																																																																																																																																	
Db	494	ADDISLLK 501																																																																																																																																																																																																																																																	
RESULT 2																																																																																																																																																																																																																																																			
<table border="1"> <thead> <tr> <th>ID</th><th>BACK-RAT</th><th>STANDARD;</th><th>PRT;</th><th>501 AA.</th></tr> </thead> <tbody> <tr> <td>AC</td><td>P36819;</td><td></td><td></td><td></td></tr> <tr> <td>DT</td><td>30-MAY-2000 (Rel. 39, Created)</td><td></td><td></td><td></td></tr> <tr> <td>DT</td><td>30-MAY-2000 (Rel. 39, Last sequence update)</td><td></td><td></td><td></td></tr> <tr> <td>DT</td><td>01-OCT-2000 (Rel. 40, Last annotation update)</td><td></td><td></td><td></td></tr> <tr> <td>DE</td><td>BETA-SECRETASE PRECURSOR (EC 3.4.23.-) (BETA-SITE APP CLEAVING ENZYME)</td><td></td><td></td><td></td></tr> <tr> <td>DE</td><td>(BETA-SITE AMYLOID PRECURSOR PROTEIN CLEAVING ENZYME) (ASPARTYL</td><td></td><td></td><td></td></tr> <tr> <td>DE</td><td>PROTEASE 2) (ASP 2) (ASP2) (MEMBRANE-ASSOCIATED ASPARTIC</td><td></td><td></td><td></td></tr> <tr> <td>DE</td><td>(MEMAPASIN-2).</td><td></td><td></td><td></td></tr> <tr> <td>GN</td><td>BACE.</td><td></td><td></td><td></td></tr> <tr> <td>OS</td><td>Rattus norvegicus (Rat).</td><td></td><td></td><td></td></tr> <tr> <td>OC</td><td>Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;</td><td></td><td></td><td></td></tr> <tr> <td>OC</td><td>Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.</td><td></td><td></td><td></td></tr> <tr> <td>OX</td><td>NCBI_TaxID=10116;</td><td></td><td></td><td></td></tr> <tr> <td>RN</td><td>[1]</td><td></td><td></td><td></td></tr> <tr> <td>RX</td><td>SEQUENCE FROM N.A.</td><td></td><td></td><td></td></tr> <tr> <td>RA</td><td>MEDLINE=20002972; PubMed=10531052;</td><td></td><td></td><td></td></tr> <tr> <td>RA</td><td>Vassar R., Bennett B.D., Babu-Khan S., Kahn S., Mendiaz E.A.,</td><td></td><td></td><td></td></tr> <tr> <td>RA</td><td>Denis P., Teplow D.B., Ross S., Amarante P., Loeloff R., Luo Y.,</td><td></td><td></td><td></td></tr> <tr> <td>RA</td><td>Fisher S., Fuller J., Edenson S., Lile J., Jarosinski M.A.,</td><td></td><td></td><td></td></tr> <tr> <td>RA</td><td>Biere A.L., Curran E., Burgess T., Louis J.-C., Collins F.,</td><td></td><td></td><td></td></tr> <tr> <td>RA</td><td>Treanor J., Rogers G., Citron M.,</td><td></td><td></td><td></td></tr> <tr> <td>RT</td><td>"Beta"-secretase cleavage of Alzheimer's amyloid precursor protein by</td><td></td><td></td><td></td></tr> <tr> <td>RT</td><td>the transmembrane aspartic protease BACE."</td><td></td><td></td><td></td></tr> <tr> <td>RL</td><td>Science 286:735-741(1999).</td><td></td><td></td><td></td></tr> <tr> <td>CC</td><td>-!- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC PROCESSING OF THE</td><td></td><td></td><td></td></tr> <tr> <td>CC</td><td>AMYLOID PRECURSOR PROTEIN (APP). CLEAVES AT THE AMINO TERMINUS OF</td><td></td><td></td><td></td></tr> <tr> <td>CC</td><td>THE A-BETA PEPTIDE SEQUENCE, BETWEEN RESIDUES 671 AND 672 OF APP,</td><td></td><td></td><td></td></tr> <tr> <td>CC</td><td>LEADS TO THE GENERATION AND EXTRACELLULAR RELEASE OF BETA-CLEAVED</td><td></td><td></td><td></td></tr> <tr> <td>CC</td><td>SOLUBLE APP, AND A CORRESPONDING CELL-ASSOCIATED CARBOXY-TERMINAL</td><td></td><td></td><td></td></tr> <tr> <td>CC</td><td>FRAGMENT WHICH IS LATER RELEASED BY GAMMA-SECRETASE (BY</td><td></td><td></td><td></td></tr> <tr> <td>CC</td><td>SIMILARITY).</td><td></td><td></td><td></td></tr> <tr> <td>CC</td><td>-!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.</td><td></td><td></td><td></td></tr> <tr> <td>CC</td><td>-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE</td><td></td><td></td><td></td></tr> <tr> <td>CC</td><td>EUKARYOTIC ASPARTYL PROTEASES FAMILY.</td><td></td><td></td><td></td></tr> <tr> <td>CC</td><td></td><td></td><td></td><td></td></tr> <tr> <td>CC</td><td>This SWISS-PROT entry is copyright. It is produced through a collaboration</td><td></td><td></td><td></td></tr> <tr> <td>CC</td><td>between the Swiss Institute of Bioinformatics and the EMBL outstation -</td><td></td><td></td><td></td></tr> <tr> <td>CC</td><td>the European Bioinformatics Institute. There are no restrictions on its</td><td></td><td></td><td></td></tr> <tr> <td>CC</td><td>use by non-profit institutions as long as its content is in no way</td><td></td><td></td><td></td></tr> <tr> <td>CC</td><td>modified and this statement is not removed. Usage by and for commercial</td><td></td><td></td><td></td></tr> <tr> <td>CC</td><td>entities requires a license agreement (See http://www.isb-sib.ch/</td><td></td><td></td><td></td></tr> <tr> <td>CC</td><td>or send an email to license@isb-sib.ch).</td><td></td><td></td><td></td></tr> <tr> <td>CC</td><td></td><td></td><td></td><td></td></tr> <tr> <td>DR</td><td>EMBL; AF190727; AAF04144.1; -</td><td></td><td></td><td></td></tr> <tr> <td>DR</td><td>InterPro; IPR001461; -</td><td></td><td></td><td></td></tr> <tr> <td>DR</td><td>Pfam; PF00026; asp; 3.</td><td></td><td></td><td></td></tr> <tr> <td>DR</td><td>PRINTS; PR00792; PEPSPIN.</td><td></td><td></td><td></td></tr> </tbody> </table>				ID	BACK-RAT	STANDARD;	PRT;	501 AA.	AC	P36819;				DT	30-MAY-2000 (Rel. 39, Created)				DT	30-MAY-2000 (Rel. 39, Last sequence update)				DT	01-OCT-2000 (Rel. 40, Last annotation update)				DE	BETA-SECRETASE PRECURSOR (EC 3.4.23.-) (BETA-SITE APP CLEAVING ENZYME)				DE	(BETA-SITE AMYLOID PRECURSOR PROTEIN CLEAVING ENZYME) (ASPARTYL				DE	PROTEASE 2) (ASP 2) (ASP2) (MEMBRANE-ASSOCIATED ASPARTIC				DE	(MEMAPASIN-2).				GN	BACE.				OS	Rattus norvegicus (Rat).				OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				OX	NCBI_TaxID=10116;				RN	[1]				RX	SEQUENCE FROM N.A.				RA	MEDLINE=20002972; PubMed=10531052;				RA	Vassar R., Bennett B.D., Babu-Khan S., Kahn S., Mendiaz E.A.,				RA	Denis P., Teplow D.B., Ross S., Amarante P., Loeloff R., Luo Y.,				RA	Fisher S., Fuller J., Edenson S., Lile J., Jarosinski M.A.,				RA	Biere A.L., Curran E., Burgess T., Louis J.-C., Collins F.,				RA	Treanor J., Rogers G., Citron M.,				RT	"Beta"-secretase cleavage of Alzheimer's amyloid precursor protein by				RT	the transmembrane aspartic protease BACE."				RL	Science 286:735-741(1999).				CC	-!- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC PROCESSING OF THE				CC	AMYLOID PRECURSOR PROTEIN (APP). CLEAVES AT THE AMINO TERMINUS OF				CC	THE A-BETA PEPTIDE SEQUENCE, BETWEEN RESIDUES 671 AND 672 OF APP,				CC	LEADS TO THE GENERATION AND EXTRACELLULAR RELEASE OF BETA-CLEAVED				CC	SOLUBLE APP, AND A CORRESPONDING CELL-ASSOCIATED CARBOXY-TERMINAL				CC	FRAGMENT WHICH IS LATER RELEASED BY GAMMA-SECRETASE (BY				CC	SIMILARITY).				CC	-!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.				CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE				CC	EUKARYOTIC ASPARTYL PROTEASES FAMILY.				CC					CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				CC	the European Bioinformatics Institute. There are no restrictions on its				CC	use by non-profit institutions as long as its content is in no way				CC	modified and this statement is not removed. Usage by and for commercial				CC	entities requires a license agreement (See http://www.isb-sib.ch/				CC	or send an email to license@isb-sib.ch).				CC					DR	EMBL; AF190727; AAF04144.1; -				DR	InterPro; IPR001461; -				DR	Pfam; PF00026; asp; 3.				DR	PRINTS; PR00792; PEPSPIN.			
ID	BACK-RAT	STANDARD;	PRT;	501 AA.																																																																																																																																																																																																																																															
AC	P36819;																																																																																																																																																																																																																																																		
DT	30-MAY-2000 (Rel. 39, Created)																																																																																																																																																																																																																																																		
DT	30-MAY-2000 (Rel. 39, Last sequence update)																																																																																																																																																																																																																																																		
DT	01-OCT-2000 (Rel. 40, Last annotation update)																																																																																																																																																																																																																																																		
DE	BETA-SECRETASE PRECURSOR (EC 3.4.23.-) (BETA-SITE APP CLEAVING ENZYME)																																																																																																																																																																																																																																																		
DE	(BETA-SITE AMYLOID PRECURSOR PROTEIN CLEAVING ENZYME) (ASPARTYL																																																																																																																																																																																																																																																		
DE	PROTEASE 2) (ASP 2) (ASP2) (MEMBRANE-ASSOCIATED ASPARTIC																																																																																																																																																																																																																																																		
DE	(MEMAPASIN-2).																																																																																																																																																																																																																																																		
GN	BACE.																																																																																																																																																																																																																																																		
OS	Rattus norvegicus (Rat).																																																																																																																																																																																																																																																		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;																																																																																																																																																																																																																																																		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.																																																																																																																																																																																																																																																		
OX	NCBI_TaxID=10116;																																																																																																																																																																																																																																																		
RN	[1]																																																																																																																																																																																																																																																		
RX	SEQUENCE FROM N.A.																																																																																																																																																																																																																																																		
RA	MEDLINE=20002972; PubMed=10531052;																																																																																																																																																																																																																																																		
RA	Vassar R., Bennett B.D., Babu-Khan S., Kahn S., Mendiaz E.A.,																																																																																																																																																																																																																																																		
RA	Denis P., Teplow D.B., Ross S., Amarante P., Loeloff R., Luo Y.,																																																																																																																																																																																																																																																		
RA	Fisher S., Fuller J., Edenson S., Lile J., Jarosinski M.A.,																																																																																																																																																																																																																																																		
RA	Biere A.L., Curran E., Burgess T., Louis J.-C., Collins F.,																																																																																																																																																																																																																																																		
RA	Treanor J., Rogers G., Citron M.,																																																																																																																																																																																																																																																		
RT	"Beta"-secretase cleavage of Alzheimer's amyloid precursor protein by																																																																																																																																																																																																																																																		
RT	the transmembrane aspartic protease BACE."																																																																																																																																																																																																																																																		
RL	Science 286:735-741(1999).																																																																																																																																																																																																																																																		
CC	-!- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC PROCESSING OF THE																																																																																																																																																																																																																																																		
CC	AMYLOID PRECURSOR PROTEIN (APP). CLEAVES AT THE AMINO TERMINUS OF																																																																																																																																																																																																																																																		
CC	THE A-BETA PEPTIDE SEQUENCE, BETWEEN RESIDUES 671 AND 672 OF APP,																																																																																																																																																																																																																																																		
CC	LEADS TO THE GENERATION AND EXTRACELLULAR RELEASE OF BETA-CLEAVED																																																																																																																																																																																																																																																		
CC	SOLUBLE APP, AND A CORRESPONDING CELL-ASSOCIATED CARBOXY-TERMINAL																																																																																																																																																																																																																																																		
CC	FRAGMENT WHICH IS LATER RELEASED BY GAMMA-SECRETASE (BY																																																																																																																																																																																																																																																		
CC	SIMILARITY).																																																																																																																																																																																																																																																		
CC	-!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.																																																																																																																																																																																																																																																		
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE																																																																																																																																																																																																																																																		
CC	EUKARYOTIC ASPARTYL PROTEASES FAMILY.																																																																																																																																																																																																																																																		
CC																																																																																																																																																																																																																																																			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration																																																																																																																																																																																																																																																		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -																																																																																																																																																																																																																																																		
CC	the European Bioinformatics Institute. There are no restrictions on its																																																																																																																																																																																																																																																		
CC	use by non-profit institutions as long as its content is in no way																																																																																																																																																																																																																																																		
CC	modified and this statement is not removed. Usage by and for commercial																																																																																																																																																																																																																																																		
CC	entities requires a license agreement (See http://www.isb-sib.ch/																																																																																																																																																																																																																																																		
CC	or send an email to license@isb-sib.ch).																																																																																																																																																																																																																																																		
CC																																																																																																																																																																																																																																																			
DR	EMBL; AF190727; AAF04144.1; -																																																																																																																																																																																																																																																		
DR	InterPro; IPR001461; -																																																																																																																																																																																																																																																		
DR	Pfam; PF00026; asp; 3.																																																																																																																																																																																																																																																		
DR	PRINTS; PR00792; PEPSPIN.																																																																																																																																																																																																																																																		


```

Db 14 SGMLPAQTHLGRPLRLSGLAGPLGLRLPRETDESEPCRRGSFVEMVDNLRGKSSQ 73
Qy 61 GYVEMTVCSPTQNLNLYDVTGSSFAVCAAPHPFLHRYQRLSTYDLRKGVVYPY 120
Db 74 GYVEMTVCSPTQNLNLYDVTGSSFAVCAAPHPFLHRYQRLSTYDLRKGVVYPY 133
Qy 121 QKWEGELGTLVSPHGPNTVVRANIAITSDKFFNGSNWEGILGLAYAEIARPD 180
Db 134 QKWEGELGTLVSPHGPNTVVRANIAITSDKFFNGSNWEGILGLAYAEIARPD 193
Qy 181 LEFFFDLSLVKQTHPNLFSQLCGAGFPUNQSEVLASVGGSMIIGDHSLYTGS 240
Db 194 LEFFFDLSLVKQTHPNLFSQLCGAGFPUNQSEVLASVGGSMIIGDHSLYTGS 253
Qy 241 IREWEYEVIIYRVEINGDLKMDCKEYNDKSIYDVGTTNLRPKKVEAAVSIK 300
Db 254 IREWEYEVIIYRVEINGDLKMDCKEYNDKSIYDVGTTNLRPKKVEAAVSIK 313
Qy 301 STEKFPDGFNLGEQLVCWAGTTPNIFPVISLYLMGEVTNQSFRITILPQOYL 360
Db 314 STEKFPDGFNLGEQLVCWAGTTPNIFPVISLYLMGEVTNQSFRITILPQOYL 373
Qy 361 ATSQDCYKFAISQSGTGTVMGAVIMEGYVYVDFRARKRIGFAVSACHVHDEF 420
Db 374 ATSQDCYKFAVSQSGTGTVMGAVIMEGYVYVDFRARKRIGFAVSACHVHDEF 433
Qy 421 PFVTLDMDCGYNIPQTDSTLMTAYMAAICAFMLPLCLMVCOWCLRLCROHDD 480
Db 434 PFVTLDMDCGYNIPQTDSTLMTAYMAAICAFMLPLCLMVCOWCLRLCROHDD 493
Qy 481 ADDISLLK 488
Db 494 ADDISLLK 501

RESULT 4
ID BAE2_HUMAN STANDARD; PRT: 518 AA.
AC Q9Y5Z0; Q9UJ76;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE BETA SECRETASE 2 PRECURSOR (EC 3.4.23.-) (BETA-SITE APP-CLEAVING
DE ENZYME 2) (ASPARTYL PROTEASE 1) (ASP 1) (ASPL1) (MEMBRANE-ASSOCIATED
DE ASPARTIC PROTEASE 1) (MEMAP SIN-1).
GN BACE2 OR ASP21.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20057170; PubMed=10591213;
RA Yan R., Bienkowski M.J., Shuck M.E., Miao H., Tory M.C., Pauley A.M.,
RA Brashier J.R., Strachan N.C., Mathews W.R., Buhl A.E., Carter D.B.,
RA Tomasselli A.G., Parodi L.A., Heinrichson R.D., Gurney M.E.;
RT "Membrane-anchored aspartyl protease with Alzheimer's disease
RT beta-secretase activity.";
RL Nature 402:533-537(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RA Xin H., Stephens J.C., Duan X., Harrowe G., Kim E., Grieshammer U.,
RA Giese K.;
RT "Identification of a novel aspartic-like protease differentially
RT expressed in human breast cancer cell lines.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Accarino M.P., Fumagalli P., Ottolenghi S., Taramelli R.;
RT "Cloning of a gene from chromosome 21 Down region encoding a potential
RT transmembrane aspartyl protease.";
```

```

RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Solans A., Estivill X., de la Luna S.;
RT "Cloning of a novel mammalian aspartyl protease.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=20030166; PubMed=10561122;
RA Hussain I., Powell D.J., Howlett D.R., Tew D.G., Meek T.D.,
RA Chapman C., Glover I.S., Murphy K.E., Southan C.D., Ryan D.M.,
RA Smith T.S., Simmons D.L., Walsh F.S., Dingwall C., Christie G.;
RT "Identification of a novel aspartic proteinase (Asp 2) as
RT beta-secretase.";
RL Mol. Cell. Neurosci. 14:419-427(1999).
RN [6]
RP SEQUENCE FROM N.A.
RX MEDLINE=20144060; PubMed=10677483;
RA Lin X., Koelsch G., Wu S., Downs D., Dashti A., Tang J.;
RT "Human aspartic protease memapsin 2 cleaves the beta-secretase site of
RT beta-amyloid precursor protein.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:1456-1460(2000).
RN [7]
RP SEQUENCE FROM N.A.
RX MEDLINE=20283799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordieck G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Rieselmann L., Dagand E.,
RA Wehmeyer S., Borzlym K., Gardiner K., Nizetic D., Francis F.,
RA Lehrach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 21.";
RL Nature 405:311-319(2000).
RN [8]
RP SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC [9]
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
RN [10]
EMBL; AF200342; AAF17078.1; -
EMBL; AF117892; AAD45240.1; -
EMBL; AF050171; AAD45963.1; -
EMBL; AF178532; AAF29494.1; -
EMBL; AF204944; AAF36368.1; -
EMBL; AF200192; AAF13714.1; -
EMBL; AF163284; CAB90458.1; -
EMBL; AL163285; CAB90554.1; -
InterPro: IPR001461; -
Pfam: PF00026; asp; 3.
PRINTS; PR00792; PEP SIN
PROSITE; PS00141; ASP_PROTEASE; 2.
KW Hydrolyase; Aspartyl protease; Glycoprotein; Zymogen; Transmembrane;
Signal.
FT SIGNAL 1 20 POTENTIAL.
FT PROPEP 21 ? POTENTIAL.
FT CHAIN ? 518 BETA SECRETASE 2.
FT DOMAIN 21 473 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 474 494 POTENTIAL.
FT DOMAIN 495 518 CYTOPLASMIC (POTENTIAL).
FT ACT_SITE 110 110 BY SIMILARITY.
FT ACT_SITE 303 303 BY SIMILARITY.
```

FT CARBOHYD 170 170 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 366 366 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 36 36 A -> T (IN REF. 6).
SQ SEQUENCE 518 AA; 56180 MW; 2E903150823760D3 CRC64;

Query Match 45.4%; Score 1175; DB 1; Length 518;
Best Local Similarity 47.1%; Pred. No. 7.5e-87;
Matches 228; Conservative 81; Mismatches 151; Indels 24; Gaps 5;

QY 5 PAHQTHGIRLPLRSGLGAPLGLRLPRETDEPEPEGRGSEVENVNLRGKSGGYV 64
DB 52 PAERHADGLALEPALA-----SPAGANFLAMVNLQDGSGRGYL 94
QY 65 EMTVGSPPQTLNLIVDTGSSNFVAGAAPFLHRYQRLSSYRDLRGVYVPTQGW 124
DB 95 EMLIGTPQKQLIVDTGSSNFVAVACTPHSYIDYEDTERSSYRSGKGFDTVKYTOGSW 154
QY 125 EGEGLTDLVSIPIHPNVTVRANIAATESDKFFINGSNIEGLGLAYAEIARDDSLPFF 184
DB 155 TGFVGEDLVTPKGFNTSELVNIATIFESENFFLPGLKWNGLGLAYATLAKPSSLET 214
QY 185 FDSLVTQTHVPLNLSLQLCGAGFLNQSEVLASVGGSMIIGDHSLYTGSLWYTPIRRE 244
DB 215 FDSLVYQANIPNVFSQMGAGLPAVGS---GTNGSLVLGGIEPSLYKGDHWYTIKEE 271
QY 245 WYVEYIIVRVEINGQDLKMDCKEYNDKSIYDSTGTTNLRPLKPKVFAAASKAASSTK 304
DB 272 WYQIETLKEIGQSLNLDREYNADKAIVDSGTTLLRLPQKVFDAVVEAVARASLIPE 331
QY 305 FPGFWLGEOLVQWAGTTPWNLFFVLSYLMGEVNTNOSFRILPQOYLRPVEDVATSO 364
DB 332 FSDGFWTGSOLACWTNSETPWSYFPKISYILRDENSRSFRILPQLYTQPMGAGLNY 391
QY 365 DDCYKFAISOSSGTGVGMVIMGEFVYVDFRARKRIGFAVSACHVDFPTAAVEGPEVT 424
DB 392 -ECYRGISPTNVALIGATVMGEFVYVDFRARKRIGFAVSPCAEATAGAVSEISGPFST 450
QY 425 LDMEDCGYNTPOQDESLMTIAYVMAAIC-ALEMLPLCLMVCQWRCURCQRHDDFADD 483
DB 451 EDVASNCVPAQSLSEPLTWISYALMSVCGAILLVLLVLLLLLPFCOR--RPRDPEVND 508
QY 484 ISLL 487
DB 509 ESSL 512

RESULT 5
PEP1_GADMO STANDARD; PRT; 324 AA.
AC P56272;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PEP1IN IIB (EC 3.4.23.-).
OS Gadus morhua (Atlantic cod).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadoidei; Gadidae;
OC Gadus.
OX NCBI_TaxID=8049;
RN [1]
RP SEQUENCE, AND X-RAY CRYSTALLOGRAPHY.
RC TISSUE=Stomach;
RA Karlsten S., Hough E., Olsen R.L.;
RT "Structure and proposed amino-acid sequence of a pepsin from Atlantic
RT cod (Gadus morhua).";
RL Acta Crystallogr. D 54:32-46(1998).
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
DR PDB; 1AM5; 24-DEC-97.
DR InterPro: IPR001461;
DR InterPro: IPR001969; -.

DR PFAM: PF00026; asp; 1.
DR PRINTS; PR00792; PEP1IN.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
KW Hydrolase; Aspartyl protease; Digestion; Stomach; Gastric juice;
KW 3D-structure.
FT ACT_SITE 32 32 BY SIMILARITY.
FT ACT_SITE 214 214 BY SIMILARITY.
FT DISULFID 45 50 BY SIMILARITY..
FT DISULFID 205 209 BY SIMILARITY.
FT DISULFID 247 280 BY SIMILARITY.
SQ SEQUENCE 324 AA; 34014 MW; EE3A6097B6941DD7 CRC64;

Query Match 12.6%; Score 327; DB 1; Length 324;
Best Local Similarity 27.9%; Pred. No. 5.3e-19;
Matches 104; Conservative 66; Mismatches 137; Indels 66; Gaps 15;

QY 50 MVDNLRGKSGQGYVEMTVGSPPTLNILVDTGSSNFVAG---AAPHPLHRYQRLS 105
DB 2 VTEQMKNEADTEYGVISIGTPPESEKVFIFDTGSSNLVSSSHCSAQACSNHKKFKPQS 61
QY 106 STYRDLRKGVYVPTQCKWEGELGTDLVSIPIHG--PNVTVRANIAATESDKFFINGSNW 163
DB 62 STYVETGKTVDLTYGTGGMRGILGQDVTSVGGSDPNQELG---ESQTEPGPQA-AAPF 117
QY 164 EGIGLAYABIAIARDDSLPFFDSLVKQTHV-PNLFSLQLCGAGFLNQSEVLASVGGSM 222
DB 118 DGLTGL-AYPSIAAA--GAVPFDNMGSQLVEKDLFSYLSGGG--ANGSEVM----- 166
QY 223 IIGSDHSLYTGSLWYTPIRREYVYIIVRVEINGQDLKMD-CKEYNYDKSIYDSTGTTN 281
DB 167 -LGGVNSHTYGSIHWPVTAEXYVQVLDGTVNGTAAACEG-----QAVDVTGTSK 219
QY 282 LRLPKVFEAAVSKAASSTKFPDGLGEOLVQWAGTTPWNLFFVLSYLMGEVNTN 341
DB 220 IVAPVSAANIMKDIGASEN-----QGEHMGN---CASVQSLPDIIF-----TI 260
QY 342 QSPFRICLPQOYLRPVEDVATSOQDCKYKFAISOSSGTG-----VGMVIMGEFVYV 393
DB 261 NGVKQPLPPSAITEGDQAFCTSD-----GLGSSGVPSNTSELWIFGVFLRNYYTII 311
QY 394 DRARKRIGFAVSA 406
DB 312 DRTNKNVGFAPAA 324

RESULT 6
PEP1_RABIT STANDARD; PRT; 387 AA.
AC P28712;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PEP1IN II-1 PRECURSOR (EC 3.4.23.1) (PEPSIN A).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91009127; PubMed=2129536;
RA Kageyama T., Tanabe K., Koizumi O.;
RT "Structure and development of rabbit pepsinogens. Stage-specific
RT zymogens, nucleotide sequences of cDNAs, molecular evolution, and
RT gene expression during development";
RL J. Biol. Chem. 265:17031-17038(1990).
CC -I- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
CC ALSO CLEAVED TO SOME EXTENT.
CC -I- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.
CC -I- DEVELOPMENTAL STAGE: PEPSINOGENS IN GROUP I, II, AND III WHERE
CC THE PREDOMINANT ZYMOGENS AT LATE POSTNATAL STAGE.
CC -I- MISCELLANEOUS: THE EXPRESSION OF PEPSINOGEN GENES IS REGULATED BY

CC	-!	HORMONES AND RELATED SUBSTANCES.
CC	-!	SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC	PIR;	B38302; B38302.
CC	HSP;	P00791; IPSA.
DR	MEROPS;	A01.001; -.
DR	InterPro;	IPR001461; -.
DR	InterPro:	IPR001969; -.
DR	Pfam;	PF00026; asp. 1.
DR	PRINTS;	PR00792; PEPsin.
DR	PROSITE;	PS00141; ASP-PROTEASE; 2.
DR	KW	Hydrolase; Aspartyl protease; Digestion; Stomach; Gastric juice; Zymogen; Signal; Phosphorylation; Multigene family.
FT	SIGNAL	1 15
FT	PROPEP	16 59
FT	CHAIN	60 387
FT	MOD_RES	129 93
FT	ACT_SITE	93 93
FT	ACT_SITE	276 276
FT	DISULFID	106 111
FT	DISULFID	267 271
FT	DISULFID	310 343
SSQ	SEQUENCE	387 AA; 42070 MW; A6EC48F71554IA48 CRC64;
Query Match	11.8%;	Score 306; DB 1; Length 387;
Best Local Similarity	27.1%;	Pred. No. 3.3e-17;
Matches	98; Conservative	67; Mismatches 131; Indels 66; Gaps 15;
QY	62	YVEMTVGSPQTNLNILDVTGGSNFAVG---AAPHPFLHRYQRQLSSTYRDLRKGVVV 117
DB	75	IETGISITGPQEFTVFIDTGGSNLWVPSTYCSLACFLHKRPDSDTSFFQATSETLSI 134
QY	118	PTYOGKWEGELGTDLVSIIPHGNVTVRANTAATESD---KFFINGSNWEGILGLAYAEI 174
DB	135	TYGCTSGMTGILLGYDPVKV---GNIEDTNQIFGLSKTEPGITFLV--APFDGILGLAYPSI 189
QY	175	ARPDSLEPPFDLSLVKQHV-PNFLSLQLCGAGPLNQSEVLASVSGSMIGGDHSLYT 233
DB	190	SASDAT--PVFDNMWNEGLVSDFSVYLSNG-----EKSVMWFSGIDSSYYT 237
QY	234	GSLWTPIRREWEYEVIVIRVEINGQDLKM--DCKEYNYSVDSTNNLRPKKVFEA 291
DB	238	GSLNWVPVSHGYQWITMDSITINGETIACADSC-----QAVVDGTSLLAGPTSISAIK 291
QY	292	AVKSKAAASTKEPFGFWLGEOLV-CWOAGTTPNWIFFPVISLYLMGEVNTNQSFRTITLP 350
DB	292	IOSYIGASKNL-----LGENIISCASIDSLPDIVF-----TINN 325
QY	351	QQYLREPVED-VATSODDC---YKFAISOSSGT--VMGAVIMEGFYVVFBRKRGRGFV 404
DB	326	VQYPPLPASAYILKEDDDCLSGFDGMNLDTSYGELMWLGDFVFIQYFTVFDNRANNQVGLAA 385
QY	405	SA 406
DB	386	AA 387
RESULT	7	
ID	PEPA_MACFU	STANDARD; PRT; 388 AA.
AC	P27678;	
DT	01-AUG-1992	(Rel. 23, Created)
DT	01-AUG-1992	(Rel. 23, Last sequence update)
DE	01-NOV-1995	(Last annotation update)
DE	PEFSIN A-4 PRECURSOR	(EC 3.4.23.1) (PEPSIN I/II).
GN	PGA.	
OC	Macaca fuscata fuscata	(Japanese macaque).
OC	Eukaryota;	Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia;	Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC	Cercopithecoidea;	Macaca.
OX	NCBL_Taxid=9543;	
RN	[1]	

RP	SEQUENCE FROM N.A., AND SEQUENCE OF 16-70.	
TISSUE-Gastric mucosa;		
RC	MEDLINE=92037645; PubMed=1935977;	
RA	"Kageyama T., Tanabe K., Koizumi O.;	
RT	"Development-dependent expression of isozymes of monkey	
RL	pepsinogens and structural differences between them.";	
RL	Eur. J. Biochem. 202:205-215(1991)	
CC	-!- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE ALSO CLEAVED TO SOME EXTENT.	
CC	-!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.	
CC	-!- MISCELLANEOUS: THE EXPRESSION OF Pepsin GENES IS REGULATED BY HORMONES AND RELATED SUBSTANCES.	
CC	-!- MISCELLANEOUS: EACH Pepsinogen IS CONVERTED TO CORRESPONDING Pepsin AT pH 2.0 IN PART AS A RESULT OF THE RELEASE OF A 47 AA ACTIVATION SEGMENT AND IN PART AS A RESULT OF STEPWISE PROTEOLYTIC CLEAVAGE VIA AN INTERMEDIATE FORM(S).	
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE EUKARYOTIC ASPARTYL PROTEASES FAMILY.	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).	
CC	EMBL; X59753; CAA42425.1; -.	
DR	PIR; S16065; S16065.	
DR	PIR; S19682; S19682.	
DR	HSP; P00790; IPSO.	
DR	MEROPS; A01.001; -.	
DR	InterPro; IPR001461; -.	
DR	InterPro:	IPR001969; -.
DR	Pfam; PF00026; asp. 1.	
DR	PRINTS; PR00792; PEPsin.	
DR	PROSITE; PS00141; ASP-PROTEASE; 2.	
DR	KW	Hydrolase; Aspartyl protease; Digestion; Stomach; Gastric juice; Zymogen; Multigene family; Signal; Glycoprotein.
KW	SIGNAL	1 15
FT	PROPEP	16 38
FT	PROPEP	39 62
FT	CHAIN	63 388
FT	ACT_SITE	94 94
FT	ACT_SITE	277 277
FT	DISULFID	107 112
FT	DISULFID	268 272
FT	DISULFID	311 344
FT	CARBOHYD	88 88
FT	SEQUENCE	388 AA; 41955 MW; A2923AB1F7CDEB9 CRC64;
SSQ	SEQUENCE	388 AA; 41955 MW; A2923AB1F7CDEB9 CRC64;
Query Match	11.8%;	Score 304.5; DB 1; Length 388;
Best Local Similarity	27.6%;	Pred. No. 4.4e-17;
Matches	108; Conservative	64; Mismatches 136; Indels 83; Gaps 17;
QY	31	PRETDERPEEPGRGSFVEMVDNLRGKSGQGYVEMTVGSPQTNLNILDVTGSSNFVGA 90
DB	60	PTLIDPOPLE-----NYLDV-----EYFGTIGTGTPAQNVTFVDFDGSSNLW-- 102
QY	91	APHPEL-----HRYQRQLSSTYRDLRKGVVVPYTOGWCELCGLDLVSIIPHGNVTV 143
DB	103	--PSVYCYSACMDHNLFPQDSSTYRATSKTSITYGTGSMTGILGYDPVKV---GGISD 158
QY	144	RANIAAIATESDK-EFFINGSNWEGILGLAYAEIARPDSDSLEPPFDLSLVKQHV-PNFLSLQ 201
DB	159	TNQIFGLSETPEGFLEYAFDPDGLGLAYPSIS--SSGATPVFDNIWNQRLVSQDLFSYV 216
QY	202	LCAAGFPLNQSEVLASVSGSMIGGDHSLYTGLSWTYPIRREWYEVIVIRVEINGQDL 261
DB	217	LSAD-----DQS-----GSSVIFGIDSSYTGSLNWVPVSVGEIWOISVDSITMGKI 266

```

QY 262 --KMDCKEYNDKSIYDSGTTNRLPKPKVFAAIVKSIKAASSTKPPDGFNLGEQLV-CW 318
DB 267 ACARKG-----QAIVDTGTSLLTGTSPITANTQSDISENSD-----GEMVYSCS 312
QY 319 QAGCTPWNIPFVLSLVLGMGEVWNSFRITLPOQY-LRPVEDVATSDDCYK-----FAI 372
DB 313 AISSLPDIVF-----TINGVOYPLPPSAIYLQSGSGCTSGFQGMVDP 354
QY 373 SQSSTGTVMGAVIMEGFYVVFDRARKRIGFA 403
DB 355 TESGELWILGDVFIROYFTVDFDRANNOVGLA 385

RESULT 8
PEPA_CHICK
ID PEPA_CHICK STANDARD; PRT; 367 AA.
AC P00793;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PEPSIN A PRECURSOR (EC 3.4.23.1).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE.
RX MEDLINE=84004412; PubMed=6617663;
RA Baudys M., Kostka V.;
RT "Covalent structure of chicken pepsinogen.";
RL Eur. J. Biochem. 136:89-99(1983).
CC -1- INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
CC ALSO CLEAVED TO SOME EXTENT.
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
DR PIR; A00984; PECH.
DR HSSP; P00794; 3CMS.
DR MEROPS; A01.001; -.
DR InterPro; IPR001461; -.
DR InterPro; IPR001969; -.
DR Pfam; PF00026; asp; 1.
DR PRINTS; PR00792; PEPsin.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
KW Hydrolase; Aspartyl protease; Digestion; Stomach; Zymogen;
KW Glycoprotein; Gastric juice.
FT PROPEP 1 42 ACTIVATION PEPTIDE.
FT CHAIN 43 367 PEPSIN A.
FT ACT_SITE 77 77
FT ACT_SITE 260 260
FT CARBOHYD 113 113 N-LINKED (GLCNAC. . .).
FT DISULFID 90 95
FT DISULFID 251 255
FT DISULFID 290 323
SQ SEQUENCE 367 AA; 40431 MW; 0C547E7FD8F5B341 CRC64;

Query Match 11.7%; Score 302; DB 1; Length 367;
Best Local Similarity 24.0%; Pred. No. 6.4e-17;
Matches 88; Conservative 69; Mismatches 125; Indels 84; Gaps 13;

QY 62 YVEMTVGSPQTLNVLDTGSSNFAYGAAPHPL-----HRYYQRLSSVYDLRKG 114
DB 59 YGCTISIGTQQDFSVIFDTGSSNLW---PSIYCKSSACSNHKKRDPKSSVTVSNET 115
QY 115 YVVPYTGKWEGLGTDLVSIPLGPNVTVRANIAATRESOK-FPINGSNWEGLGLAYAE 173
DB 116 VYIAYTGSMGILGYDTAV---SSIDVQNIQFGLSETEPGSPFYCNFDGILGLAFPS 172
QY 174 IARPDSDLEPFDFSLVQTHV-PNLFSLQLCGAGFFPLNQSEVLASVCGSMILGIDHSLY 232

```

```

DB 173 IS--SSGATPVFDNMMSQHLVAQDLFSVYLSKDG-----ETGSFVLFGLDPNT 220
QY 233 TGSLSWYTPIRREYYEIVVRVEINGDCLK--MDCKEYNDKSIYDSGTTNRLPKPKVFE 290
DB 221 TKGIYVWPLSAEYIWOITMDRVTGNGKYVACFFTC-----QAIVDTGTSLLVMPQAYN 274
QY 291 AAVKSKAKASSTE-----KFPDGFNLGEQLVWAGTTPWNIPFVLSLVLGMGEVWNS 343
DB 275 RIILKGLVSSDGEISCDISKLDP-----VTFHNGHA----- 307
QY 344 FRITILPQOYLRPVEDVATSDDCYKFAISQSSTGT-----VNGAVIMEGFYVVFDRAR 397
DB 308 -----FTLPASAVLVNEDGSCMLGFENMGTPTELGEQWILGDVFIROYVIFEDRAN 358
QY 398 KRIGFA 403
DB 359 NKVGLS 364

RESULT 9
CATEL_HUMAN
ID CATEL_HUMAN STANDARD; PRT; 396 AA.
AC P14031;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CATHEPSIN E PRECURSOR (EC 3.4.23.34).
GN CTSE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89380302; PubMed=2674141;
RA Azuma T., Pals G., Mohandas T.K., Couvreur J.M., Taggart R.T.;
RT "Human gastric cathepsin E. Predicted sequence, localization to
RT chromosome 1, and sequence homology with other aspartic
RT proteinases.";
RL J. Biol. Chem. 264:16748-16753(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92112877; PubMed=1370478;
RA Azuma T., Liu W.G., Vander Laan D.J., Bowcock A.M., Taggart R.T.;
RT "Human gastric cathepsin E gene. Multiple transcripts result from
RT alternative polyadenylation of the primary transcripts of a single
RT gene locus at 1q31-q32.";
RL J. Biol. Chem. 267:1609-1614(1992).
RN [3]
RP SEQUENCE OF 54-68; 77-95; 141-155; 275-285 AND 389-396.
RX MEDLINE=90241267; PubMed=2334440;
RA Athauda S.B.P., Matsuzaki O., Kageyama T., Takahashi K.;
RT "Structural evidence for two isozymic forms and the carbohydrate
RT attachment site of human gastric cathepsin E.";
RL Biochem. Biophys. Res. Commun. 168:878-885(1990).
CC -1- FUNCTION: DUE TO ITS INTRACELLULAR LOCATION AND DISTRIBUTION IN
CC LYMPHOID ASSOCIATED TISSUE, IT MAY HAVE A ROLE IN IMMUNE FUNCTION.
CC -1- CATALYTIC ACTIVITY: SIMILAR TO CATHEPSIN D, BUT SLIGHTLY BROADER
CC SPECIFICITY.
CC -1- SUBUNIT: HOMODIMER; LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M84424; AAA52300.1; -.
DR EMBL; M84413; AAA52300.1; JOINED.

```

```

DR EMBL; M84417; AAA52300.1; JOINED.
DR EMBL; M84418; AAA52300.1; JOINED.
DR EMBL; M84419; AAA52300.1; JOINED.
DR EMBL; M84420; AAA52300.1; JOINED.
DR EMBL; M84421; AAA52300.1; JOINED.
DR EMBL; M84422; AAA52300.1; JOINED.
DR EMBL; J05036; AAA52300.1; -.
DR PIR; A34401; A34401.
DR PIR; A34643; A34643.
DR PIR; A42038; A42038.
DR HSP; P00794; 3CMS.
DR MEROPS; A01.010; -.
DR MIM; I16890; -.
DR InterPro; IPR001461; -.
DR InterPro; IPR001969; -.
DR Pfam; PF00026; asp; 1.
DR PRINTS; PR00792; PEPsin.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
KW Hydrolase; Aspartyl protease; Glycoprotein; Zymogen; Signal.
FT SIGNAL 1 17
FT PROPEP 18 53
FT CHAIN 54 396
FT MOD_RES 18 18
FT ACT_SITE 96 96
FT ACT_SITE 281 281
FT DISULFID 60 60
FT DISULFID 109 114
FT DISULFID 272 276
FT DISULFID 314 351
FT CARBOHYD 90 90
FT CARBOHYD 220 220
FT CARBOHYD 333 333
FT SEQUENCE 396 AA; 42793 MW; 40B643C5FB01521E CRC64;

Query Match 11.6%; Score 299.5; DB 1; Length 396;
Best Local Similarity 25.9%; Pred. No. 1.1e-16;
Matches 100; Conservative 65; Mismatches 148; Indels 73; Gaps 16;

QY 35 DEPEEPRGRGSEVMDNLRGSGQGYVEMTVGSPQTLNLTVDGSSNFAVGA--- 90
DB 63 DOSAKEP-----LNYLD-----MEYFTGISGPPQNTVFIDTGSNLMWPSVYCT 110

QY 91 APHPFLHRYQRLSSYRDLRGVVPYTGQKWEGLGTLVSIPIHPGNVTVRANIAAI 150
DB 111 SPACKTHSRQPSOSSYSPQGSFSIQYGTGSLGIAGADQVSV-EGLTVVQQFGESV 169

QY 151 TESDKFFINGSNWEGLGLAYAEIARPDSDLSEPFDSLVKQTHVPNLFSLQLCGAGFPLN 210
DB 170 TEPGQTEVD-AEFDGILGLGYPSLA--VGGVTPVDFNMMAQ-----NLVDLPMSFVYMSSN 222

QY 211 QSEVLASVGGSMIIGGDHSLYGLSWTYTPIRREWYEVIIIVRVEINGQDLKMDCKEYNY 270
DB 223 PE---GGAGSELIFGGYDHSFGSLNMWVPTQAWQIADLNQVGG--TVMFCSE--G 275

QY 271 DKSTVDGTTNLRPKRVFAEAKVSKAASSTKFPDGEWLGQVLCVQAGATTPWNIFFV 330
DB 276 QCAIVDGTSLITGSPDKIKQLQNAICAA-----VDGEYAVE-----CANLNVMPD 322

QY 331 ISLYLMEGVNQSPRITILPQOYLPRVDEIVATSDQDCYKFAISQSTG----- 378
DB 323 VTFTING-----VPYTLSPATY--TLDFVDGMQFC-----SSGFGOLDIHPAGP 366

QY 379 -TVNGVIMEGVYVDFRARRKRIGFA 403
DB 367 LWILGDVFIQFYVDFRGNRVGLA 392

RESULT 10
PEPE_CHKCK
ID PEPE_CHKCK STANDARD; PRT; 383 AA.
AC P16476;
DT 01-AUG-1990 (Rel. 15, Created)

```

```

DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE EMBRYONIC PEPSINOGEN PRECURSOR (EC 3.4.23.-).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88227903; Pubmed=3131317;
RA Hayashi K., Agata K., Mochii M., Yasugi S., Eguchi G., Mizuno T.;
RT "Molecular cloning and the nucleotide sequence of cDNA for embryonic
RT chicken pepsinogen: phylogenetic relationship with prothymosin.";
RL J. Biochem. 103:290-296(1988).
CC -!- DEVELOPMENTAL STAGE: SPECIFICALLY SECRETED DURING THE EMBRYONIC
CC PERIOD IN THE CHICKEN PROVENTRICULUS (GLANDULAR STOMACH).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D00215; BAA00153.1; -.
DR PIR; A41443; A41443.
DR HSP; P00794; 3CMS.
DR MEROPS; A01.028; -.
DR InterPro; IPR001461; -.
DR InterPro; IPR001969; -.
DR Pfam; PF00026; asp; 1.
DR PRINTS; PR00792; PEPsin.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
DR Hydrolase; Aspartyl protease; Digestion; Signal; Glycoprotein.
FT SIGNAL 1 16
FT CHAIN 17 383
FT ACT_SITE 94 94
FT ACT_SITE 276 276
FT DISULFID 107 112
FT DISULFID 267 271
FT DISULFID 310 344
FT CARBOHYD 132 132
FT CARBOHYD 204 204
FT CARBOHYD 309 309
FT CARBOHYD 350 350
FT VARIANT 51 51
FT SEQUENCE 383 AA; 41719 MW; 1642796871611F54 CRC64;

Query Match 11.5%; Score 298.5; DB 1; Length 383;
Best Local Similarity 25.2%; Pred. No. 1.3e-16;
Matches 90; Conservative 75; Mismatches 125; Indels 67; Gaps 14;

QY 62 YXVETVSGPPQTLNLTVDGSSNFAVGA---APHPFLHRYQRLSSYRDLRGVTV 117
DB 76 YXGTISIGTPPDFTVFTDGTSSNLMWPSVCTSPACQSHQMFNPSSQSTYKSTQNLST 135

QY 118 PTQKWEGLGTLVSIPIHPGNVTVRANIAAITESDKFFINGSNWEGLGLAYAEIARP 177
DB 136 HYGTGMEGTGVCDDTVTVASLMDTNQLFGLST-SEPGQFFVY-VKPDGILGLGYPSLA- 192

QY 178 DQSLPEFFDSLVKQTHV-PNLFSLQLCGAGFPLNQSEVLASVGGSMIIGGDHSLYGL 236
DB 193 -DGITVFDNMWNSLEQLNLSVYLS-----REPMSVVFSGIDESYFTGSI 240

QY 237 WYTPIRREWYEVIIIVRVEINGQDL--KMDCKEYNDKSIDVSGTTNLRPKRVFAEAKV 294
DB 241 NWIPVSYQGYWQISMDSIIVNKGIEACSSGC-----QAIDTGTSLVAGPASDINDIOS 294

```


QY 295 SIKASSTKFFDGFGLVQWAGTTPNFFPVISL-----YLMGEVNTOSFRITILP 350
DB 295 AVGANQI-----YGEYSV-----NCSHILAMPDVVFIIGGI----- 326
QY 351 QOYLPRVEDVA-----TSQDDCKVFAISQSSTGTVMGAVIMEGFYVDFDRARKRIGFA 403
DB 327 -QY--PVPALAYTEONGQCTCMTSSSFQNSADLWILGDVFIRVYYSIFDRANRNVGLA 380

RESULT 11
CATD_HUMAN
ID CATD_HUMAN STANDARD; PRT: 412 AA.
AC P07339;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CATHEPSIN D PRECURSOR (EC 3.4.23.5).
GN CTSD.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85270436; PubMed=3927292;
RA Faust P.L., Kornfeld S., Chirgwin J.M.;
RT "Cloning and sequence analysis of cDNA for human cathepsin D.";
RN [2]
RP proc. Natl. Acad. Sci. U.S.A. 82:4910-4914(1985).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=87231068; PubMed=3588310;
RA Westley B.R., May F.E.B.;
RT "Oestrogen regulates cathepsin D mRNA levels in oestrogen responsive
human breast cancer cells";
RN [4]
RP Nucleic Acids Res. 15:3773-3786(1987).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=91299158; PubMed=2069717;
RA Redeker B., Heckendorf B., Grosch H.W., Mersmann G., Hasilik A.;
RT "Molecular organization of the human cathepsin D gene";
RN [6]
RP DNA Cell Biol. 10:423-431(1991).
RN [7]
RP SEQUENCE OF 1-22 FROM N.A.
RX MEDLINE=94085791; PubMed=8426386;
RA May F.E., Smith D.J., Westley B.R.;
RT "The human cathepsin D-encoding gene is transcribed from an estrogen-
regulated and a constitutive start point.";
RN [8]
RP Gene 134:277-282(1993).
RN [9]
RP SEQUENCE OF 1-22 FROM N.A.
RX MEDLINE=95021301; PubMed=7935485;
RA Augereau P., Miralles F., Cavaillès V., Gaudelot C., Parker M.,
Rocheport H.;
RT "Characterization of the proximal estrogen-responsive element of
human cathepsin D gene";
RN [10]
RP Mol. Endocrinol. 8:693-703(1994).
RN [11]
RP SEQUENCE OF 170-180.
RN [12]
RP TISSUE=Liver;
RA Hochstrasser D.F., Frutiger S., Paquet N., Bairoch A., Ravier F.,
Pasquali C., Sanchez J.-C., Tissot J.-D., Bjellqvist B., Vargas R.,
Appel R.D., Hughes G.J.;
RL Submitted (JUN-1992) to the SWISS-PROT data bank.
RN [13]
RP X-RAY CRYSTALLOGRAPHY (3 ANGSTROMS).
RX MEDLINE=93223670; PubMed=8467789;
RA Metcalf P., Fusek M.;
RT "Two crystal structures for cathepsin D: the lysosomal targeting
signal and active site";
RN [14]
RP EMBO J. 12:1293-1302(1993).
RN [15]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE=93342076; PubMed=8393577;

RA Baldwin E.T., Bhat T.N., Gulnik S., Hosur M.V., Sowder R.C. II,
Cachau R.E., Collins J., Silva A.M., Erickson J.W.;
RT "Crystal structures of native and inhibited forms of human cathepsin
D: implications for lysosomal targeting and drug design.";
RN [16]
RP Proc. Natl. Acad. Sci. U.S.A. 90:6796-6800(1993).
RN [17]
RP FUNCTION: CATHEPSIN D IS AN ACID PROTEASE ACTIVE IN INTRACELLULAR
PROTEIN BREAKDOWNS.
RN [18]
RP SUBUNIT: CONSISTS OF A LIGHT CHAIN AND A HEAVY CHAIN.
RN [19]
RP SUBCELLULAR LOCATION: LYSOSOMAL.
RN [20]
RP SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
EUKARYOTIC ASPARTYL PROTEASES FAMILY.
RN [21]
RP This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by not-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
RN [22]
RP EMBL: M11233; AAB59529.1; -
RN [23]
RP EMBL: X05344; CAA28955.1; -
RN [24]
RP EMBL: M63138; AAA51922.1; -
RN [25]
RP EMBL: M63134; AAA51922.1; JOINED.
RN [26]
RP EMBL: M63135; AAA51922.1; JOINED.
RN [27]
RP EMBL: M63136; AAA51922.1; JOINED.
RN [28]
RP EMBL: M63137; AAA51922.1; JOINED.
RN [29]
RP EMBL: L12980; AAA16314.1; -
RN [30]
RP EMBL: S74689; AAD14156.1; -
RN [31]
RP EMBL: S52557; AAD13868.1; -
RN [32]
RP PIR: A25771; KHHUD.
RN [33]
RP PDB: 1LYA; 31-JAN-94.
RN [34]
RP PDB: 1LYB; 31-JAN-94.
RN [35]
RP MEROPS: A01.009; -
RN [36]
RP SWISS-2DPAGE; P07339; HUMAN.
RN [37]
RP MIM: 116840; -
RN [38]
RP InterPro: IPR001461; -
RN [39]
RP InterPro: IPR001969; -
RN [40]
RP Pfam: PF00026; asp; 1.
RN [41]
RP PRINTS: PR00792; PEPsin.
RN [42]
RP PROSITE: PS00141; ASP_PROTEASE; 2.
RN [43]
KW Hydrolase; Aspartyl protease; Glycoprotein; Lysosome; Signal; Zymogen;
RN [44]
KW 3D-structure.
RN [45]
FT SIGNAL 1 18
FT PROPEP 19 64 ACTIVATION PEPTIDE.
FT CHAIN 65 412 CATHEPSIN D.
FT CHAIN 65 161 LIGHT CHAIN (PROBABLE).
FT CHAIN 169 412 HEAVY CHAIN (PROBABLE).
FT ACT_SITE 97 97
FT ACT_SITE 295 295
FT DISULFID 91 160
FT DISULFID 110 117
FT DISULFID 286 290
FT DISULFID 329 366
FT CARBOHYD 134 134
FT CARBOHYD 263 263 N-LINKED (GLCNAC. . .).
FT STRAND 67 74 N-LINKED (GLCNAC. . .).
FT TURN 75 77
FT TURN 78 85
FT TURN 86 89
FT STRAND 90 97
FT TURN 98 99
FT STRAND 103 107
FT TURN 108 109
FT TURN 112 113
FT TURN 115 118
FT TURN 119 119
FT STRAND 123 123
FT TURN 125 127
FT TURN 129 130
FT STRAND 132 141
FT STRAND 146 158
FT STRAND 172 184

FT HELIX 188 192
 FT STRAND 197 200
 FT HELIX 204 206
 FT HELIX 208 210
 FT HELIX 214 220
 FT TURN 221 222
 FT STRAND 228 233
 FT STRAND 243 247
 FT TURN 248 248
 FT STRAND 252 254
 FT STRAND 255 263
 FT STRAND 267 267
 FT TURN 268 269
 FT STRAND 270 279
 FT TURN 280 281
 FT STRAND 284 285
 FT TURN 287 288
 FT STRAND 290 294
 FT TURN 296 297
 FT STRAND 298 298
 FT STRAND 301 303
 FT HELIX 305 315
 FT TURN 316 316
 FT STRAND 318 319
 FT TURN 322 323
 FT STRAND 325 328
 FT HELIX 329 334
 FT STRAND 338 342
 FT TURN 343 344
 FT STRAND 345 349
 FT HELIX 351 354
 FT STRAND 355 357
 FT TURN 359 362
 FT STRAND 365 368
 FT STRAND 370 372
 FT TURN 377 379
 FT STRAND 383 385
 FT HELIX 387 390
 FT TURN 391 392
 FT STRAND 393 398
 FT TURN 399 402
 FT STRAND 403 409
 SQ SEQUENCE 412 AA; 44552 MW; 903FBB8412E0CF0B0 CRC64;

Query Match 11.5%; Score 296.5; DB 1; Length 412;
 Best Local Similarity 26.8%; Pred. No. 2.1e-16;
 Matches 117; Conservative 66; Mismatches 163; Indels 91; Gaps 19;

QY 13 IRLPLR-----SGLGGAPLGL-----RLPRETDEPEPEPGRGSGFVEMWD 52
 DB 22 VRIPLHKFTSIRRTSEVGGSGVEDLIAGPVSKYSQAVPAVTE-----GPIPEVLK 72
 QY 53 NLRGSGGOGYVEMTVGPPQTLNVLVDGTSSNFAVGAAPHFL-----HRYVQROLSS 106
 DB 73 NYMDAQ---YGEIGIGTPPQCFTVFTDGTSSNLWVPSIHCKLLDIACWIHHKNSDKSS 129
 QY 107 TYRDLRKGVPYPTOCKWEGELGDLVSIPI-----HGPNTVVRANIAAITESDKFF 157
 DB 130 TVVKNGTDFDIHYGSGSLGSLVSDQTVSPQCQSASSALAGVKVQRVGFGEATKQPGIT 189
 QY 158 INGSNWEGILGLAVAEIARPDSDLEPPFDLSLVKQTHV-PNLFSLQLCGAGFPLNQSEVLA 216
 DB 190 FIAAKFDGILGNAPRIS--VNNVLPFDNLWQKLVQDNIFSYL-----SRDPDA 239
 QY 217 SVGGSMITGGIDHSLYTGSLNVTPIRREWEYVIVRVEI-NGQDLKMDCKEYNDKSIIV 275
 DB 240 QPGGELMLGGTDSKYKGLSYLWTRKAYQVHLDQVEVASGLTL--CKE--GCEAIV 294
 QY 276 DSGTNTNLRPKKVEAAVKSITKAASSTKPEPDGFWLGEQLV-CWQAGTTWNIPFVVISLY 334
 DB 295 DTGTSVMWGPVDEVELQKAGVPLIQ-----GEYMIPCEKYST-----LPAITLK 341

QY 335 LMGEVNTQSFRTILPQOYLRPVEDVATSDDCYKFAISQ-----SSTGTVNGAVIMEGF 389
 DB 342 LGG---KGYKLS--PEDYTLKVSQAGKTL--CLSGFMGMDIPPPSGPLWILGDFVIGRY 393
 QY 390 YVDFDRARRKRIGFAVSA 406
 DB 394 YTVFDRDNRNVGFAEAA 410

RESULT 12
 PEP2_RABIT
 ID PEP2_RABIT STANDARD; PRT; 387 AA.
 AC P27821;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE PEP2IN II-2/3 PRECURSOR (EC 3.4.23.1) (PEPSIN A).
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91009127; PubMed=2129536;
 RA Kageyama T., Tanabe K., Koiwai O.;
 RT "Structure and development of rabbit pepsinogens. Stage-specific
 RT zymogens, nucleotide sequences of cDNAs, molecular evolution, and
 RT gene expression during development.";
 RL J. Biol. Chem. 265:17031-17038(1990).
 CC -!- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
 CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
 CC ALSO CLEAVED TO SOME EXTENT.
 CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.
 CC -!- DEVELOPMENTAL STAGE: PEPSINOGENS IN GROUP I, II, AND III WHERE
 CC THE PREDOMINANT ZYMOGENS AT LATE POSTNATAL STAGE.
 CC -!- MISCELLANEOUS: THE EXPRESSION OF PEPSINOGEN GENES IS REGULATED BY
 CC HORMONES AND RELATED SUBSTANCES.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
 CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----

EMBL; M59235; AAA85369.1;
 PIR; C38302; C38302.
 HSP; P00790; LP50.
 MEROPS; A01.001;
 InterPro; IPR001461;
 InterPro; IPR001969;
 Pfam; PF00026; asp. 1.
 PRINTS; PR00792; PEP2IN.
 DR PROSITE; PS00141; ASP_PROTEASE; 2.
 DR Hydrolyase; Aspartyl protease; Digestion; Stomach; Gastric juice;
 KW Zymogen; Signal; Phosphorylation; Multigene family.
 FT SIGNAL 1 15
 FT PROPEP 16 59 ACTIVATION PEPTIDE.
 FT CHAIN 60 387 PEPSIN II-2/3.
 FT MOD_RES 129 129 PHOSPHORYLATION (POTENTIAL).
 FT ACT_SITE 93 93 BY SIMILARITY.
 FT ACT_SITE 276 276 BY SIMILARITY.
 FT DISULFID 105 111 BY SIMILARITY.
 FT DISULFID 267 271 BY SIMILARITY.
 FT DISULFID 310 343 BY SIMILARITY.
 SQ SEQUENCE 387 AA; 42100 MW; 66FC331A3DC75891 CRC64;

Query Match 11.4%; Score 296; DB 1; Length 387;
 Best Local Similarity 26.9%; Pred. No. 2.1e-16;

Matches 97; Conservative 63; Mismatches 135; Indels 66; Gaps 13;

```
QY 62 YVEMTGPSPQTLNLTVDGSSNFAVGAHPHF-----LHRYYQRLSSTYRDLRKG 114
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 75 YFGTISIGTPQDFVIFDTGSSNLWV---PSTYCSLACALHKKRFNEDSDSYTGTSSET 131
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 115 VVVPYTGKWEGLGELTDLVSPHGPNNVVRANIAAITESDKFFINGSNWEGILGLAYABI 174
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 132 LSTYTGTSMTGILGYDVTVKVGSIEDTNIQFLGSKTEPSLTFLF--APDGLGLAYPSI 189
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 175 ARPDSLEPFDLSLVKQTHV-PNLESLOLCGAGFPNLQNSVLASVGGSMILGIDHSYLT 233
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 190 SSSDAT--PVFNMMNNEGLVSQDLFSVLYSSDD-----ERKSLVMFGGIDSSYYT 237
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 234 GSLWTPPIRREWYEVIIIVRVEINGQDLKM--DCKEYNDKSIDVSGTTLNLRPKKVFEA 291
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 238 GSLNWPVSVYEGYQWITMDSVSINGETINACDSC-----QAIVDTGTSLLTGP---TS 287
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 292 AVKSKIAASSTKFPDGFGLWGLQV-CWQAGTTPNIPVVISLYLMGEVNTNOSFRITILP 350
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 288 AISNIQSYTGASK-----NLGENVISCSAIDSLPDIVF-----TING 325
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 351 QOYLAPVEDVATSDQDCKYKFAISQSSTGP-----VMGAVIMEGVVVFDRARKRIGRAV 404
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 326 IQYPLPASAYILKEDDDCTSGLEGHNVDTYTGELWILGDVFIHQYFTVFDRAANNQLGLAA 385
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 405 S 405
Db 386 A 386
```

RESULT 13

```
ID CATE_CAVPO STANDARD; PRT; 391 AA.
AC P25796;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE CATHEPSIN E PRECURSOR (EC 3.4.23.34).
GN CTSE.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX NCBI_TxID=10141;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=92355614; PubMed=1644829;
RA Kageyama T., Ichinose M., Tsukada S., Miki K., Kurokawa K., Koiwai O.,
RA Tanji M., Yakabe E., Athuda S.B., Takahashi K.;
RT "Gastric procathepsin E and progastricsin from guinea pig.
RT Purification, molecular cloning of cDNAs, and characterization of
RT enzymatic properties, with special reference to procathepsin E.";
RL J. Biol. Chem. 267:16450-16459(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Gastric mucosa;
RX MEDLINE=96073637; PubMed=8540321;
RA Kageyama T., Ichinose M., Miki K., Moriyama A., Yonezawa S.,
RA Tanji M., Athuda S.B., Takahashi K.;
RT "Isolation, characterization, and structure of procathepsin E and
RT cathepsin E from the gastric mucosa of guinea pig.";
RL Adv. Exp. Med. Biol. 362:211-221(1995).
CC -I- FUNCTION: DUE TO ITS INTRACELLULAR LOCATION AND DISTRIBUTION IN
CC LYMPHOID ASSOCIATED TISSUE, IT MAY HAVE A ROLE IN IMMUNE FUNCTION.
CC -I- CATALYTIC ACTIVITY: SIMILAR TO CATHEPSIN D, BUT SLIGHTLY BROADER
CC SPECIFICITY.
CC -I- SUBUNIT: HOMODIMER; LINKED BY A DISULFIDE BOND.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

```
DR EMBL; M88653; AAA37052.1; -
DR EMBL; S80347; AAB35844.1; -
DR PIR; A43356; A43356.
DR HSSP; P00794; 3CMS.
DR MEROPS; A01.010; -.
DR InterPro; IPR001461; -.
DR InterPro; IPR001969; -.
DR Pfam; PF00026; asp; 1.
DR PRINTS; PR00792; PEPsin.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
KW Hydrolase; Aspartyl protease; Glycoprotein; Zymogen; Signal.
FT SIGNAL 1 18
FT PROPEP 19 53
FT CHAIN 54 391
FT ACT_SITE 92 92
FT ACT_SITE 276 276
FT DISULFID 56 56
FT DISULFID 105 110
FT DISULFID 267 271
FT DISULFID 309 346
FT CARBOHYD 86 86
SQ SEQUENCE 391 AA; 42132 MW; 78D216BF8CFDABD CRC64;
```

Query Match 11.4%; Score 296; DB 1; Length 391;
Best Local Similarity 26.9%; Pred. No. 2.1e-16;
Matches 98; Conservative 64; Mismatches 130; Indels 72; Gaps 16;

```
QY 62 YVEMTGPSPQTLNLTVDGSSNFAVGA-----APHPFLHRYYQRLSSTYRDLRKGVVV 117
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 74 YFGTISIGTPQDFVIFDTGSSNLWVPSVCTSPACOTHPVPHPSLSSTYREVGNSESI 133
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 118 PYTGKWEGLGELTDLVSPHGPNNVVRANIAAITESDKFFINGSNWEGILGLAYABIAR 177
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 134 QYGTGSLTGIIGADQVSV-EGLTVVGQQFGESVQEPGKTFVH-AEFDGILGLGYPSLAA- 190
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 178 DDSLPEPFDLSLVKQTHVHPNLFSLOLCGAGFPNLQNSVLASVGGSMILGIDHSYLTGSLW 237
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 191 -GVPTVPFVNMAQ-----NLVALPM-----FSVMSSNPGSGSELTFGGYDPSHFSGSLN 241
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 238 YTPIRREWYEVIIIVRVEINGQDLKMDCKEYNDKSIDVSGTTLNLRPKKVFEAAVKSK 297
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 242 WVPVTQAYWQIALDGIQVG--DSVMFCSE--GCQAIVDTGTSLLTGP-----PGKIKQLQ 293
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 298 AASSTKFPDGFGLWGLQVLCWQAGTTPNIPVVISLYLMGEVNTNOSFRI-----TILPOQ 352
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 294 EALGATYVDEGY---SVQC-----ANLNMMLDVT---FIINGVPTTLNPTA 333
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 353 YLRPVEDVATSDQDCKYKFAISQSSTG-----TVMGAVIMEGVVVFDRARKR 399
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 334 Y--TLLDVFDGMOVC-----STGFEGLEIQPPAGPLWILGDVFIHQYFVAFDRGNR 383
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 400 IGFA 403
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 384 VGLA 387
```

RESULT 14

```
PEP4_RABIT
ID PEP4_RABIT STANDARD; PRT; 387 AA.
AC P28713;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PEPsin II-4 PRECURSOR (EC 3.4.23.1) (PEPSIN A).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:37:36 ; Search time 231.42 Seconds
(without alignments)
278.994 Million cell updates/sec

Title: US-09-603-713-2
Perfect score: 2587
Sequence: 1 AGVLPAGTQHGIRLPLRSG.....CLRLRQOHDFADDISLLK 488

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_16.*

- 1: sp-archaea.*
- 2: sp-bacteria.*
- 3: sp-fungi.*
- 4: sp-human.*
- 5: sp-invertebrate.*
- 6: sp-mammal.*
- 7: sp-mhc.*
- 8: sp-organelle.*
- 9: sp-phage.*
- 10: sp-plant.*
- 11: sp-rodent.*
- 12: sp-unclassified.*
- 13: sp-vertebrate.*
- 14: sp-virus.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Match	Length	ID	Description
1	2036	78.7	396	Q9ULS1	Q9ulsl1 homo sapien
2	1160.5	44.9	439	Q9H2V8	Q9h2v8 homo sapien
3	1155	44.6	514	Q9JLV18	Q9jlv18 mus musculus
4	976	37.7	468	Q9NZL2	Q9nzl2 homo sapien
5	971	37.5	396	Q9NZL1	Q9nzl1 homo sapien
6	712.5	27.5	213	Q9PD2	Q9pd2 homo sapien
7	596.5	23.1	255	Q9R1P7	Q9rip7 mus musculus
8	332.5	12.9	391	Q9VXK6	Q9vxk6 drosophila
9	332	12.8	354	Q9GYX7	Q9gyx7 boophilus m
10	309	11.9	384	Q9DEC2	Q9dec2 xenopus lae
11	308	11.9	385	Q9DEC4	Q9dec4 rana catesb
12	307	11.9	346	Q9TS27	Q9ts27 bos taurus
13	307	11.9	372	Q9VLK3	Q9vln3 drosophila
14	305	11.8	386	Q9GMY7	Q9gmy7 rhinolophus
15	305	11.8	387	Q9GMY8	Q9gmy8 sorex ungui
16	304.5	11.8	383	Q9DEC3	Q9dec3 xenopus lae
17	304.5	11.8	387	Q9DDV5	Q9ddv5 salvelinus
18	304	11.8	387	Q9GMY9	Q9gmy9 suncus muri
19	303.5	11.7	383	Q9DE45	Q9de45 salvelinus

20	302.5	11.7	376	13	Q9PUR8	Q9pur8 pleuronecte
21	302	11.7	384	13	Q91322	Q91322 rana catesb
22	301	11.6	382	13	Q9PRG9	Q9prg9 gallus gall
23	301	11.6	423	5	Q9VKP7	Q9vvp7 drosophila
24	296.5	11.5	396	13	Q93428	Q93428 chionodraco
25	295.5	11.4	386	6	Q9GMY6	Q9gmy6 canis famli
26	292	11.3	399	13	Q93458	Q93458 podarcis sl
27	290.5	11.2	381	6	Q9GK11	Q9gk11 camelus dro
28	289.5	11.2	399	13	Q9DD89	Q9dd89 brachydanio
29	287.5	11.1	444	5	Q21966	Q21966 caenorhabdi
30	284	11.0	398	13	P87370	P87370 oncorhynch
31	284	11.0	427	5	P91802	P91802 schistosoma
32	283.5	11.0	378	13	Q9PUR9	Q9pur9 pleuronecte
33	283.5	11.0	380	6	Q28950	Q28950 sus scrofa
34	283	10.9	422	5	Q96906	Q96906 onchocerca
35	281	10.9	389	13	Q9PWK1	Q9pwk1 gallus gall
36	280.5	10.8	396	13	Q9DEX3	Q9dex3 clupea hare
37	279.5	10.8	390	6	Q9GK10	Q9gk10 camelus dro
38	278	10.7	370	6	Q9TTW1	Q9ttw1 bos taurus
39	278	10.7	387	6	Q9N2D4	Q9n2d4 callithrix
40	278	10.7	389	13	Q9W643	Q9w643 gallus gall
41	277	10.7	446	5	Q9N9H3	Q9n9h3 necator ame
42	275	10.6	385	6	Q29080	Q29080 sus scrofa
43	275	10.6	413	3	Q14413	Q14413 pichia angu
44	273.5	10.6	410	5	Q9VQ13	Q9vql3 drosophila
45	273	10.6	388	6	Q9GMY2	Q9gmy2 oryctolagus

ALIGNMENTS

RESULT 1
Q9ULS1 PRELIMINARY; PRT; 396 AA.
AC Q9ULS1
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE KIAA1149 'PROTEIN (FRAGMENT).'
GN KIAA1149.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=20039618; PubMed=10574461;
RA Hirose M., Nagase T., Ishikawa K., Kikuno R., Nomura N., Ohara O.;
RT "Characterization of cDNA clones selected by the Genemark analysis
from size-fractionated cDNA libraries from human brain.";
RL DNA Res. 6:329-336(1999).
DR EMBL; AB032975; BAA86463.1; -.
DR HSSP; P56272; 1AM5
DR InterPro; IPR001461; -.
DR Pfam; PF00026; asp; 2.
DR PRINTS; PR00792; PEPsin.
FT NON_TER 1
SQ SEQUENCE 396 AA; 44428 MW; A2CBCD52DCC089E0 CRC64;

Query Match 78.7%; Score 2036; DB 4; Length 396;
Best Local Similarity 100.0%; Pred. No. 2.7e-157;
Matches 384; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	105	SSYRDLRGVYVYTGKWEGLTDLVSIPIHPGNVTVRANIAATITKDFKFNINGSNW	164
Db	13	SSYRDLRGVYVYTGKWEGLTDLVSIPIHPGNVTVRANIAATITKDFKFNINGSNW	72
QY	165	GILGLAYAEIARPDSDSLEPFDSLVKQTHVPLNLSQLCAGFPPLNQSEVLASVGGSMII	224
Db	73	GILGLAYAEIARPDSDSLEPFDSLVKQTHVPLNLSQLCAGFPPLNQSEVLASVGGSMII	132

Db	357	IAGA	AVSEISG	PSTED	VASNCVPAQ	SLSEP	ILWIVSYALMSV	CGAILLVLLVLL	LPFR	416
QY	469	CL	CL	RO	QHD	DF	AD	DIS	LL	487
Db	417	Q	R	--	RPR	DE	VN	DE	SS	433
RESULT	3									
ID	Q9JL18	PRELIMINARY; PRT; 514 AA.								
AC	Q9JL18									
DT	01-OCT-2000	(TrEMBLrel. 15, Created)								
DT	01-OCT-2000	(TrEMBLrel. 15, Last sequence update)								
DT	01-MAR-2001	(TrEMBLrel. 16, Last annotation update)								
DE	ASPARTYL	PROTEASE 1.								
OS	Mus musculus	(Mouse).								
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;									
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.									
OX	NCBI_TaxID=10090;									
RN	[1]									
RP	SEQUENCE FROM N.A.									
RA	Choi D.K., Sugano S., Sakaki Y.;									
RT	"Molecular characterization of the mouse Aspl gene, a homolog of the									
RL	human Aspl (Down Syndrome Region aspartyl protease).";									
DR	Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.									
DR	EMBL; AF216310; AAF36599.1; -									
DR	InterPro; IPR001461; -									
DR	Pfam; PF00026; asp; 3.									
DR	PRINTS; PR00792; PEPSIN.									
DR	PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_2.									
KW	Protease.									
SQ	SEQUENCE 514 AA; 55799 MW; A70725F2C1DF5B47 CRC64;									
Query Match	44.6%;	Score	1155;	DB	11;	Length	514;			
Best Local Similarity	48.5%;	Pred. No.	1.6e-85;							
Matches	225;	Conservative	76;	Mismatches	143;	Indels	20;	Gaps	5;	
QY	1	AG	VL	PA	HG	TQ	HG	I	R	L
Db	43	AS	AV	P	G	L	G	T	P	--
QY	61	G	Y	V	E	M	T	G	S	P
Db	87	G	Y	L	E	M	L	I	G	T
QY	121	O	K	W	E	G	E	L	G	T
Db	147	Q	S	W	T	G	F	V	E	D
QY	181	L	E	P	F	D	S	L	V	K
Db	207	L	E	T	F	D	S	L	V	A
QY	241	I	R	R	E	Y	E	V	I	V
Db	264	I	K	E	E	Y	I	E	I	L
QY	301	S	T	E	K	P	D	G	F	W
Db	324	L	I	P	E	S	D	G	F	W
QY	361	A	T	S	O	D	C	Y	K	F
Db	384	G	N	Y	-	E	C	R	F	G
QY	421	P	F	V	T	L	D	M	E	C
Db	443	P	F	T	E	D	I	A	S	N
RESULT	4									


```
Q9NZL2
ID Q9NZL2 PRELIMINARY; PRT: 468 AA.
AC Q9NZL2;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE ASPARTYL PROTEASE.
GN BACE2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Solans A., Estivill X., de la Luna S.;
RT "Cloning of a novel mammalian aspartyl protease.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF188276; AAF35835.1; -.
DR InterPro: IPR001461; -.
DR Pfam: PF00026; asp. 3.
DR PRINTS: PR00792; PEPsin.
DR PROSITE: PS00141; ASP_PROTEASE; UNKNOWN_2.
KW Protease.
SQ SEQUENCE 468 AA; 50324 MW; 717E0920126A0142 CRC64;

Query Match 37.7%; Score 976; DB 4; Length 468;
Best Local Similarity 41.1%; Pred. No. 4.9e-71;
Matches 199; Conservative 75; Mismatches 136; Indels 74; Gaps 6;

QY 5 PAHGTHGIRLPLRSLGGLGAPLGLRLPRETDEPEEPRGRGSGFVEMVDNLRKSGGGYV 64
  || || || || || || || || || || || || || || || || || || || || || ||
Db 52 PAERHADGLALALEPALA-----SPAGAAFLAMVDNLOGDSGRGYL 94

QY 65 EMTVSGPPOTLNLVDGTSSNFAVGAAPHPFLHRYQROLSTYRDLRGVVPYTOGKW 124
  || || || || || || || || || || || || || || || || || || || || || ||
Db 95 EMLGTPPKQLQLLVDGTSSNFAVAGTSHYIDTYDTERSTYRSKGFVDVTKYTOGWS 154

QY 125 EGEELGDLVSIPIHGPNTVVRANITAEKDFKFFINGSNNEGILGLAYAEIARPDSDLEPF 184
  || || || || || || || || || || || || || || || || || || || || || ||
Db 155 TGFVGEDLVTPKGFNTSLVNIATIFESNFFLPKIKWNGILGLAYATLAKPSSLET 214

QY 185 FDSLVKQTHVPLNLSLQLCAGFPLNQSEVLASVGGSMIIGDHSLSYTCGLWYTPIRRE 244
  || || || || || || || || || || || || || || || || || || || || || ||
Db 215 FDSLVQTANIPNVFSQMCGALPVAGS---GTNGGSLVLGGIEPSLYKGDWYTPIKEE 271

QY 245 WYVEVTVIIRVEINGQDLKMDCKEYNDKSIYDVGSTTNLRLPKKVFPAAYKSIKAASSTK 304
  || || || || || || || || || || || || || || || || || || || || || ||
Db 272 WYQIEILKLEIGGQSLNLDCKEYNADKAIYDSGTTLLRLPKQKVFDAVVEAVARASLL 329

QY 305 FPDGFWLGEQVLCVQAGTTPWNIFPVISLYLMGEVNTQSFRTILPQOYLRPVEDVATSQ 364
  || || || || || || || || || || || || || || || || || || || || || ||
Db 330 -----VIQPMMGAGLNV 341

QY 365 DDCYKFAISQSTGTNGAVIMEGFYVFDRAKRIGFAVSACHVIDEFRTAAVEGPFVT 424
  :||: || || || || || || || || || || || || || || || || || || || || ||
Db 342 -ECYREGISPTNALVIGATVEGFYVFDRAKRVGFAAASPCAEIAGAAVSEISGPFST 400

QY 425 LQWEDCGYINPQDSTLTIAVMAAIC-ALFMPLPLCLMVQWCRCLRCLROQDDFAD 483
  || || || || || || || || || || || || || || || || || || || || || ||
Db 401 EDVASCNCPAQSLSEPIILVSVLMSVCGAIIIVLIVLLLPFRQCR--RPRDPEVND 458

QY 484 ISLL 487
  ||
Db 459 ESSL 462

RESULT 5
Q9NZL1
ID Q9NZL1 PRELIMINARY; PRT: 396 AA.
AC Q9NZL1;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
```

```
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE ASPARTYL PROTEASE.
GN BACE2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Solans A., Estivill X., de la Luna S.;
RT "Cloning of a novel mammalian aspartyl protease.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF188277; AAF35836.1; -.
DR InterPro: IPR001461; -.
DR Pfam: PF00026; asp. 2.
DR PRINTS: PR00792; PEPsin.
DR PROSITE: PS00141; ASP_PROTEASE; UNKNOWN_2.
KW Protease.
SQ SEQUENCE 396 AA; 43013 MW; 5023A7AF391CEAC9 CRC64;

Query Match 37.5%; Score 971; DB 4; Length 396;
Best Local Similarity 50.9%; Pred. No. 9.7e-71;
Matches 189; Conservative 55; Mismatches 97; Indels 30; Gaps 5;

QY 5 PAHGTHGIRLPLRSLGGLGAPLGLRLPRETDEPEEPRGRGSGFVEMVDNLRKSGGGYV 64
  || || || || || || || || || || || || || || || || || || || || || ||
Db 52 PAERHADGLALALEPALA-----SPAGAAFLAMVDNLOGDSGRGYL 94

QY 65 EMTVSGPPOTLNLVDGTSSNFAVGAAPHPFLHRYQROLSTYRDLRGVVPYTOGKW 124
  || || || || || || || || || || || || || || || || || || || || || ||
Db 95 EMLGTPPKQLQLLVDGTSSNFAVAGTSHYIDTYDTERSTYRSKGFVDVTKYTOGWS 154

QY 125 EGEELGDLVSIPIHGPNTVVRANITAEKDFKFFINGSNNEGILGLAYAEIARPDSDLEPF 184
  || || || || || || || || || || || || || || || || || || || || || ||
Db 155 TGFVGEDLVTPKGFNTSLVNIATIFESNFFLPKIKWNGILGLAYATLAKPSSLET 214

QY 185 FDSLVKQTHVPLNLSLQLCAGFPLNQSEVLASVGGSMIIGDHSLSYTCGLWYTPIRRE 244
  || || || || || || || || || || || || || || || || || || || || || ||
Db 215 FDSLVQTANIPNVFSQMCGALPVAGS---GTNGGSLVLGGIEPSLYKGDWYTPIKEE 271

QY 245 WYVEVTVIIRVEINGQDLKMDCKEYNDKSIYDVGSTTNLRLPKKVFPAAYKSIKAASSTK 304
  || || || || || || || || || || || || || || || || || || || || || ||
Db 272 WYQIEILKLEIGGQSLNLDCKEYNADKAIYDSGTTLLRLPKQKVFDAVVEAVARASLL 331

QY 305 FPDGFWLGEQVLCVQAGTTPWNIFPVISLYLMGEVNTQSFRTILPQOYLRPVEDVATSQ 364
  || || || || || || || || || || || || || || || || || || || || || ||
Db 332 FSDGFWGSQLACWTNSETPNWYFPAKISYLRDENSRSFRITILPQK-LRVLQ----- 384

QY 365 DDCYKF-AISQ 374
  || || ||
Db 385 --CLKFPGLSQ 393

RESULT 6
Q9P0D2
ID Q9P0D2 PRELIMINARY; PRT: 213 AA.
AC Q9P0D2;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE HSPC104 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=CORD BLOOD;
RA Zhang Q.H., Ye M., Zhou J., Shen Y., Wu X.Y., Guan Z.O., Wang L.,
RA Fan H.Y., Mao Y.F., Dai M., Huang Q.H., Chen S.J., Chen Z.;
```

RT	"Human partial CDS cloned from cd34+ stem cells.";	
RL	Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AF161367; AAF28927.1; -	
DR	InterPro; IPR001461; -	
DR	Pfam; PF00026; asp; 1.	
FT	NON_TER	
SQ	SEQUENCE 213 AA; 24338 MW; EC9D3FA31CFA835C CRC64;	
	Query Match 27.5%; Score 712.5; DB 4; Length 213;	
	Best Local Similarity 83.5%; Pred. No. 3.8e-50;	
	Matches 137; Conservative 4; Mismatches 12; Indels 11; Gaps 1;	
QY	225 GGIDSLYGTSLWYTPIRREWEYVYIIVRVEINGDGLKMDCKEYNDKSIYDVGTTNLR 284	
DB	1 GGIDSLYGTSLWYTPIRREWEYVYIIVRVEINGDGLKMDCKEYNDKSIYDVGTTNLR 60	
QY	285 PKVFEAAVKSIAKASSTKFPDGEWLGEOIVCWOAGTTPWNIFPVISILYLMGEVTSNF 344	
DB	61 PKVFEAAVKSIAKASSTKFPDGEWLGEOIVCWOAGTTPWNIFPVISILYLMGEVTSNF 120	
QY	345 RITILPQQLRPVEDVATSDQDCYFAISQSSTGTVMGAVIMEG 388	
DB	121 RITILPQQLRP-----WKMPRPKTTVTVCHLTVIHG 153	
RESULT	7	
Q9R1P7	PRELIMINARY; PRT; 255 AA.	
AC	Q9R1P7;	
DT	01-MAY-2000 (TrEMBLrel. 13, Created)	
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)	
DT	01-JUN-2000 (TrEMBLrel. 14, Last annotation update)	
DE	ASPARTYL PROTEASE (FRAGMENT).	
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
OX	NCBI_TaxID=10090;	
RN	SEQUENCE FROM N.A.	
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,	
RA	Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,	
RA	George R.A., Lewis S., Richards S., Ashburner M., Henderson S.N.,	
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,	
RA	Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,	
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,	
RA	Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,	
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,	
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,	
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,	
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,	
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,	
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,	
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,	
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,	
RA	Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,	
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,	
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,	
RA	Hostal M., Houston K.A., Howland T.J., Ke Z., Kennison J.A., Ketchum K.A.,	
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,	
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Liang Y., Lin X.,	
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., McPherson D.,	
RA	Liu X., Mattel B., McIntosh T.C., McLeod M.P., Morris J., Moshrefi A.,	
RA	Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,	
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,	
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,	
RA	Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,	
RA	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,	
RA	Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,	
RA	Spieler E., Spradling A.C., Stapleton M., Strong R., Sun E.,	
RA	Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,	
RA	Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,	
RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,	
RA	Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhou S., Zhu S., Smith H.O.,	
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou S., Zhu S., Smith H.O.,	
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;	
RT	"The genome sequence of Drosophila melanogaster.";	
RL	Science 287:2185-2195(2000).	
CC	!- SIMILARITY: THE PROTEASE BELONGS TO THE PEPTIDASE FAMILY A1; ALSO	
	KNOWN AS THE EUKARYOTIC ASPARTYL PROTEASES FAMILY.	
DR	EMBL; AF003630; AAF53016.1; -	
DR	HSP; P00794; 4CMS	
DR	Flybase; FBgn0032304; CG17134.	
DR	InterPro; IPR001461; -	
DR	InterPro; IPR001969; -	
DR	Pfam; PF00026; asp; 1.	
DR	PRINTS; PR00792; PEPIN.	
DR	PROSITE; PS00141; ASP_PROTEASE; 2.	
KW	Aspartyl protease; Hydrolase.	
SQ	SEQUENCE 391 AA; 42016 MW; 99A18E7131025E1B CRC64;	

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissensbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.:
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003621; AAF52686.1; --
DR HSSP: P00797; 2REN.
DR FlyBase: FBgn0032049; CG13095.
DR InterPro: IPR001461; --
DR InterPro: IPR001969; --
DR Pfam: PF00026; asp; 1.
DR PRINTS: PR00792; PEPsin.
DR PROSITE: PS00141; ASP_PROTEASE; UNKNOWN 2.
SQ SEQUENCE 372 AA; 40080 MW; D45469E8AD72FCCE CRC64;

Query Match 11.9%; Score 307; DB 5; Length 372;
Best Local Similarity 26.0%; Pred. No. 8.1e-17;
Matches 101; Conservative 55; Mismatches 151; Indels 82; Gaps 12;

QY 29 RLP--RETDEEPEEGRGSRFVEMWDLNCKSGQGYVEMTVGSPQTNILVDTGSSNF 86
DB 49 QLPFLRSVDEE-----QLSNSNMAYIGAISITGPAQSPKVLFDGSSNL 93
QY 87 AYCA-----APHPFLHRYQRLSSTYRDLRKGYVPYTOGKWEGLGTDLVSTPHGPNVT 142
DB 94 WYPSNTCKSDACLTHNOYDSSASTVANGESFSIQYGTGSLTGYLSTDTVDV-NGLSIQ 152
QY 143 VRANTAATIESDKFFINGSNWEIGILGLAYAEIARPDSDLPEPFDLSLVKQTHVPN-LFSLQ 201
DB 153 SQTFAESTNEPCTNF-NDAFDFGILGMAYESLA--VDGVAPFPFYNMVSQGLVDNSVFSEY 209
QY 202 LCGAGFLNQSEVLASVGSGMIIGDHSGLYTGSLWYTPPIRREWEYVEIIVRVEINGODL 261
DB 210 LARDG-----TSMGCELIFGSDASLYSGALTYVPISQGYWQFTMAGSSIDGYSL 261
QY 262 KMDCKEYNDKSIYVSDGTTNLRP-----KKVFEAAVKSIIKAASSTKFFDGFHGLGQ 314
DB 262 CDDC-----QAIDTGTSLIYVNAVYITLSEILNVGEGDYLDGSSVSLPD----- 308
QY 315 LVCWQAGTTPWNIFFVISLYLNGEVTNQSFRTILPQOYLVRPVEDVATSDQCYKFAISQ 374
DB 309 -VTFNIGGTNF-----VLKFSAYIIQSGNCSAFETYM 340
QY 375 SSTGTWGVAVIMEGFYVDFDRKRIGFA 403
DB 341 GTDFWILGDFVIGQYVTEFDLGNLRIGFA 369

RESULT 14
Q9GMV7

ID Q9GMV7 PRELIMINARY; PRT; 386 AA.
AC Q9GMV7;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE PEPSINOGEN A.
GN PGNA.
OS Rhinolophus ferrumequinum.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Chiroptera; Microchiroptera; Rhinolophidae;
OC Rhinolophus.
OX NCBI_TaxID=59479;
RN [1]
RP SEQUENCE FROM N.A.
RA Narita Y., Oda S., Takenaka O., Kageyama T.;
RT "Phylogenetic position of Insectivora inferred from the cDNA sequences
of pepsinogen A and C.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB047245; BAB11750.1; --
SQ SEQUENCE 386 AA; 41591 MW; 917EE04D3166C3A4 CRC64;
Query Match 11.8%; Score 305; DB 6; Length 386;
Best Local Similarity 29.1%; Pred. No. 1.2e-16;
Matches 104; Conservative 60; Mismatches 130; Indels 64; Gaps 16;
QY 62 YVEMTVGSPQTNILVDTGSSNFAVG----AAPHPFLHRYQRLSSTYRDLRKGVYV 117
DB 74 YFTGIGTGPPOEFTVIFDTGSSNLWVPSYCSPACSNRNRENPOQSTYQGTNKLVS 133
QY 118 PYTQGWEGELGTLVSIHPGNVTVRANIAAITESEDK-FFINGSNWEIGILGLAYAEIAR 176
DB 134 AYGTGSMTGILGYDTVQV---GGITDNTQIFGLSETPGSGFLYAPDFDGLGLAYPSIA- 189
QY 177 PDSLSPEFDSLKVOPHV--PNLFLSLQLCGAGPPLNOSVLASVSGSMII-GGIDHSLYTG 234
DB 190 -SSGATPVFDNINQGLVSDQLFSVYLS-----NDQGSVWFGGIDSSYFTG 237
QY 235 SLWYTPIRREWEYVEIIVRVEINGODLKM--DCKEYNDKSIYVSDGTTNLRPKKVFEEA 292
DB 238 NLNWVFLSSTYQWITVDSTINMGQVIACSGSC-----QAIVDTCTSLLSGPTNAI-AS 290
QY 293 VYSIKAASTERKPDGFWLGEQLV-CWQAGTTPWNIFFVISLYLNGEVTNQSFRTILPQ 351
DB 291 IQGYIGASQAN-----GEMVYSCAINTLPNIYF-----TINGV 325
QY 352 QY-LREVEDVATSDDC---YKFAISQSTGT--VMGAVIMEGFYVDFDRKRIGFA 403
DB 326 QYPLPFSAYVLSQSQCGCTSGFGMDIPTSSGELWILGDFIROYFTYVDFGRNQVGLA 383
RESULT 15
Q9GMV8
ID Q9GMV8 PRELIMINARY; PRT; 387 AA.
AC Q9GMV8;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE PEPSINOGEN A.
GN PGNA.
OS Sorex unguiculatus (Long-clawed shrew).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Insectivora; Soricidae; Sorex.
OX NCBI_TaxID=62275;
RN [1]
RP SEQUENCE FROM N.A.
RA Narita Y., Oda S., Takenaka O., Kageyama T.;
RT "Phylogenetic position of Insectivora inferred from the cDNA sequences
of pepsinogen A and C.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB047244; BAB11750.1; --
SQ SEQUENCE 387 AA; 41514 MW; F2EB2E331FAA24BF CRC64;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:43:25 ; Search time 225.25 Seconds
(without alignments)
135.378 Million cell updates/sec

Title: US-09-603-713-3

Perfect score: 2663

Sequence: 1 MASMTGGQMGRCMGAGVLP.....CLRLRQHQHDFADDISLLK 503

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_0601.*

```
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT:*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT:*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT:*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT:*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT:*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT:*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT:*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT:*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT:*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT:*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT:*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT:*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT:*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT:*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT:*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT:*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT:*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT:*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT:*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT:*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT:*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2663	100.0	503	22 AAB66573	Human pro-memapsin
2	2663	100.0	503	22 AAB61335	T7 promoter and ve
3	2587	97.1	488	22 AAB66572	Human memapsin 2
4	2587	97.1	488	22 AAB61334	Memapsin 2 protein
5	2587	97.1	501	21 AAY94767	Human beta-secretase
6	2587	97.1	501	21 AAB07896	Amino acid sequenc
7	2582	97.0	501	21 AAY88425	Human aspartyl pro
8	2581	96.9	501	19 AAW59807	Amino acid sequenc
9	2519	94.6	501	21 AAY94769	Rat beta-secretase
10	2517	94.5	501	21 AAY94768	Murine beta-secret
11	2517	94.5	501	21 AAY88427	Murine aspartyl pr

12	2432.5	91.3	476	21	AAV88426	Human aspartyl pro
13	2419	90.8	456	21	AAV07897	Active enzyme port
14	2336.5	87.7	446	21	AAV88431	T7-caspase-human-p
15	2335	87.7	459	21	AAV88432	T7-caspase-human-p
16	2328	87.4	453	21	AAV88438	Modified human asp
17	2328	87.4	459	21	AAV88439	Modified human asp
18	2324	87.3	460	21	AAV07898	Amino acid sequenc
19	2320	87.1	790	19	AAV59808	Partial amino acid
20	2280	85.6	433	21	AAV88433	Human pro-Asp-2(a)
21	2156	81.0	415	21	AAV07899	Amino acid sequenc
22	2148	80.7	425	21	AAV88437	Human Asp2 amino a
23	1673	62.8	351	20	AAV35918	Extended human sec
24	1175	44.1	518	19	AAW61362	Aspartic proteinas
25	1175	44.1	518	20	AAV41714	Human PRO852 prote
26	1175	44.1	518	20	AAV22239	Human CSP56, aspar
27	1175	44.1	518	20	AAV13799	Human aspartyl pro
28	1175	44.1	518	21	AAV44270	Human PRO852 (UNQ4
29	1175	44.1	518	21	AAV88424	Human aspartyl pro
30	1077.5	40.5	423	22	AAV88479	Human membrane or
31	296.5	11.1	412	16	AAV74207	Human death associ
32	296.5	11.1	412	19	AAV71369	Death associated p
33	296.5	11.1	412	20	AAV06478	Human tumour-associ
34	296.5	11.1	412	21	AAV93685	Amino acid sequenc
35	287.5	10.8	381	13	AAV20730	Prochymosin (ptore
36	282.5	10.6	326	22	AAV66589	Human pepsin. Hom
37	282.5	10.6	326	22	AAV61351	Pepsin protein. H
38	281	10.6	545	20	AAV33830	Oleusin-spacer-Met
39	278	10.4	391	20	AAV32056	Bovine pregnancy a
40	277.5	10.4	365	4	AAV30603	Sequence encoded b
41	277.5	10.4	365	11	AAV05080	Sequence of calf p
42	277.5	10.4	375	5	AAV40078	Sequence encoded b
43	277.5	10.4	380	3	AAV20038	Pre-prorennin-A pr
44	277.5	10.4	381	5	AAV40559	Sequence of a poly
45	276.5	10.4	381	5	AAV40218	Sequence of rennin

ALIGNMENTS

RESULT 1

AAV66573
ID AAB66573 standard; Protein; 503 AA.

AC AAB66573;

DT 12-APR-2001 (first entry)

XX Human pro-memapsin 2.

DE Human; memapsin 2; nootropic; neuroprotective; amyloid precursor protein;
KW APP; memapsin 2 inhibitor; Alzheimer's disease; ss.

OS Homo sapiens.

PN WO200100665-A2.

XX 04-JAN-2001.

PF 27-JUN-2000; 2000WO-US17742.

PR 28-JUN-1999; 99US-0141363.

PR 30-NOV-1999; 99US-0168060.

PR 25-JAN-2000; 2000US-0177836.

PR 27-JAN-2000; 2000US-0178368.

PR 08-JUN-2000; 2000US-0210292.

PA (OKLA-) OKLAHOMA MEDICAL RES FOUND.

PA (UNII-) UNIV ILLINOIS FOUND.

PI Tang JUN, Hong L, Ghosh AK;

XX WPI; 2001-137933/14.

PT Novel memapsin 2 inhibitors which bind to active site of memapsin 2
PT having 2 catalytic aspartic residues and substrate binding cleft, used
PT to treat Alzheimer's disease by blocking amyloid precursor protein
PT cleavage

XX Example 4; Fig 1; 86pp; English.

XX The present sequence is given in a specification relating to an inhibitor
XX of catalytically active memapsin 2. The inhibitor binds to the memapsin 2
XX active site, which is defined by the presence of two catalytic aspartic
XX residues and a substrate binding cleft. The inhibitor is useful for
XX the treatment and diagnosis of Alzheimer's disease. It is useful in
XX screens for individuals with a genetic predisposition to Alzheimer's
XX disease. The inhibitor is useful as a reagent for specifically binding to
XX memapsin 2 or memapsin 2 analogues and for aiding in memapsin 2
XX isolation, purification and characterisation.

XX Sequence 503 AA;

Query Match 100.0%; Score 2663; DB 22; Length 503;
Best Local Similarity 100.0%; Pred. No. 9.8e-263;
Matches 503; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASMTGGQMGSGMAGVLPAGHTQHGIRLPLRSGLGAPLGLRLPRETDEPEPEGRG 60
DB 1 masmtggqmgsgmagvlpaghtqhgirplrslggagplglrlpretdeepeegrrg 60
QY 61 SFVEMVDNLRGKSGQGYVEMTVGSPPTNLILVDTGSSNFAGVGAAPFLHRYVQRQLS 120
DB 61 sfvemvndlrgksgqgyvemtvgspptnlilvdtgssnfavgaapflhryvqrqls 120
QY 121 STYRDLRGVVPYTOGKWEGLGTDLSIPHGNVTVRANIAAITSDESKFFINGSNWEG 180
DB 121 styrdlrgvvyptogkwegelgtldvsiphgnvtvraniaaitesdkffingsnweg 180
QY 181 ILGLAYAEIARPDSDLEPFFDSLVKQTHVFNLSLQLCGAGFPLNQSEVLASVGSMTIG 240
DB 181 ilglayaeiarpdsllepffdslvkqthvfnlsqldcgagfplnqsevlavsgsmiig 240
QY 241 GDHSLYTGSLWYTPTRREWYVEIIVRVEINGQDLKMDCKEYNDKSIIVDSGTTNLRIP 300
DB 241 gidhsltygslwytprrewyveiiivrveingqdlkmdckeynydksivdsgttnlrip 300
QY 301 KKVFEAAVKSIIKAASSTKFPDGFVLGEQLVCWQAGTTPWNIPFVVISLYLMGEVTSNQSFR 360
DB 301 kkvfeaaavksikaasstekfpdgfwlgeqlvcwqagttppwnifpvvislylmgevtngsfr 360
QY 361 ITILPQOYLRPVEDVATSDQCYKFAISQSTGTVMGAVIMEGFYVVFDRARKRIGFAVS 420
DB 361 itilpqylrpvedvatsdqcykfaissgstgtvmgavimegyvfvfdrarkrigfavs 420
QY 421 ACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDDESTLMTIAYVMAAICALFMLPLCLMVC 480
DB 421 achvhdefrtaavegpfvtldmedcgyinipqdestlmtiayvmaaaicalfmlplclmvc 480

RESULT 2
AAB61335
ID AAB61335 standard; protein; 503 AA.

XX AC AAB61335;

XX DT 02-APR-2001 (first entry)

XX DE T7 promoter and vector sequence.

XX KW Memapsin 2; catalyst; Alzheimer's.

XX

OS Homo sapiens.
OS Synthetic.
XX WO200100663-A2.
XX 04-JAN-2001.
XX 27-JUN-2000; 2000WO-US17661.
XX 28-JUN-1999; 99US-0141363.
XX 30-NOV-1999; 99US-0168060.
XX 25-JAN-2000; 2000US-0177836.
XX 27-JAN-2000; 2000US-0178368.
XX 08-JUN-2000; 2000US-0210292.
XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.
XX PA
XX PI Tang JJN, Lin X, Koelsch G;
XX WI; 2001-102885/11.
XX Purified recombinant catalytically active memapsin 2, used to screen
XX inhibitors of it, which are used to treat and prevent Alzheimer's
XX disease -
XX Disclosure; Fig 1; 86pp; English.
XX The present invention relates to a purified recombinant
XX catalytically active memapsin 2. The invention may be used for
XX isolating inhibitors which are used to treat or prevent
XX Alzheimer's disease. The invention may also be used to screen
XX for individuals more genetically prone to develop Alzheimer's
XX disease.
XX Sequence 503 AA;

Query Match 100.0%; Score 2663; DB 22; Length 503;
Best Local Similarity 100.0%; Pred. No. 9.8e-263;
Matches 503; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASMTGGQMGSGMAGVLPAGHTQHGIRLPLRSGLGAPLGLRLPRETDEPEPEGRG 60
DB 1 masmtggqmgsgmagvlpaghtqhgirplrslggagplglrlpretdeepeegrrg 60
QY 61 SFVEMVDNLRGKSGQGYVEMTVGSPPTNLILVDTGSSNFAGVGAAPFLHRYVQRQLS 120
DB 61 sfvemvndlrgksgqgyvemtvgspptnlilvdtgssnfavgaapflhryvqrqls 120
QY 121 STYRDLRGVVPYTOGKWEGLGTDLSIPHGNVTVRANIAAITSDESKFFINGSNWEG 180
DB 121 styrdlrgvvyptogkwegelgtldvsiphgnvtvraniaaitesdkffingsnweg 180
QY 181 ILGLAYAEIARPDSDLEPFFDSLVKQTHVFNLSLQLCGAGFPLNQSEVLASVGSMTIG 240
DB 181 ilglayaeiarpdsllepffdslvkqthvfnlsqldcgagfplnqsevlavsgsmiig 240
QY 241 GDHSLYTGSLWYTPTRREWYVEIIVRVEINGQDLKMDCKEYNDKSIIVDSGTTNLRIP 300
DB 241 gidhsltygslwytprrewyveiiivrveingqdlkmdckeynydksivdsgttnlrip 300
QY 301 KKVFEAAVKSIIKAASSTKFPDGFVLGEQLVCWQAGTTPWNIPFVVISLYLMGEVTSNQSFR 360
DB 301 kkvfeaaavksikaasstekfpdgfwlgeqlvcwqagttppwnifpvvislylmgevtngsfr 360
QY 361 ITILPQOYLRPVEDVATSDQCYKFAISQSTGTVMGAVIMEGFYVVFDRARKRIGFAVS 420
DB 361 itilpqylrpvedvatsdqcykfaissgstgtvmgavimegyvfvfdrarkrigfavs 420
QY 421 ACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDDESTLMTIAYVMAAICALFMLPLCLMVC 480
DB 421 achvhdefrtaavegpfvtldmedcgyinipqdestlmtiayvmaaaicalfmlplclmvc 480

Oy 481 QWRCLRLRQHHDPADDISLLK 503
 |||
 Db 481 gwrclrlrqghddfaddisllk 503

RESULT 3

AAB66572 3
 ID AAB66572 standard; Protein: 488 AA.

XX AC AAB66572;
 XX

DT 12-APR-2001 (first entry)
 XX

DE Human memapsin 2.
 XX

KW Human: memapsin 2; neurotropic; neuroprotective; amyloid precursor protein;
 KW APP; memapsin 2 inhibitor; Alzheimer's disease.
 XX

OS Homo sapiens.
 XX

PN WO200100665-A2.
 XX

PD 04-JAN-2001.
 XX

PF 27-JUN-2000; 2000WO-US17742.
 XX

PR 28-JUN-1999; 99US-0141363.
 PR 30-NOV-1999; 99US-0168060.
 PR 25-JAN-2000; 2000US-0177836.
 PR 27-JAN-2000; 2000US-0178368.
 PR 08-JUN-2000; 2000US-0210292.
 XX

XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.
 PA

PA (UNIT) UNIV ILLINOIS FOUND.
 XX

PI Tang JUN, Hong L, Ghosh AK;
 XX

XX WPI; 2001-137933/14.
 DR N-PSDB; AAF31848.
 XX

XX Novel memapsin 2 inhibitors which bind to active site of memapsin 2
 PT having 2 catalytic aspartic residues and substrate binding cleft, used
 PT to treat Alzheimer's disease by blocking amyloid precursor protein
 PT cleavage -
 XX

PS Example 1; Page 72-74; 86pp; English.
 XX

XX The present sequence is given in a specification relating to an inhibitor
 CC of catalytically active memapsin 2. The inhibitor binds to the memapsin 2
 CC active site, which is defined by the presence of two catalytic aspartic
 CC residues and a substrate binding cleft. The inhibitor is useful for
 CC the treatment and diagnosis of Alzheimer's disease. It is useful in
 CC screens for individuals with a genetic predisposition to Alzheimer's
 CC disease. The inhibitor is useful as a reagent for specifically binding to
 CC memapsin 2 or memapsin 2 analogues and for aiding in memapsin 2
 CC isolation, purification and characterisation.
 XX

SQ Sequence 488 AA;

Query Match 97.1%; Score 2587; DB 22; Length 488;
 Best Local Similarity 100.0%; Pred. No. 5.3e-255;
 Matches 488; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 16 AGVLPARGTOHGIRLPLRSLGGLGAPLGLRLPRETDEPEEPCRRGSFVEMVDNLKSGQ 75

Db 1 agvlpargtqhgrlplrgslrglglrplretdeepeeprrgsfvmvndnlrgksqg 60

Oy 76 GYVEMTGSPPQTNLILVDTCSSFAVGAAPHFLHRYRQLSTYRDLRKGIVVPYT 135

Db 61 gyyvemtgsppqtnilvdtgssnfavgaaphflhryyqqlsystydlrkgyvpyt 120

Oy 136 QKWEGELGTLVSLPHGPNVTVRANIAAITESDKFFINGSNWEGILGLAYAEIARPDOS 195

Db 121 gqkweigeigtldvslphgpnvtvrانياaitesdkffingsnwegilglayaelarpdds 180
 Oy 196 LEPPFDLSLVKQTHVFNLSLQLCGAGFPPLNQSEVLASVGGSMIIGGIDHSLYTGSLWYTP 255
 |||
 Db 181 lepfidslvkqthvfnlsfqlcgaagfplnqsevlavggsmilggidhslytgslyt 240
 Oy 256 IRREWYEVIIVRVEINGODLKMDCKEYNKYSDSGTTNLRPKKVFEEAAVKSIKAAS 315
 |||
 Db 241 lrrewyeviivrveinggdikmdckeynydksivdsgttnlrlpkvfeaaavksikaas 300
 Oy 316 STEKEPDGFWLGEQLVCHQACTTPHNIFFVISLYLMGEVTVNOSFRITLPPQYLRPVEDV 375
 |||
 Db 301 stekfpdgfwlgeqlvchqagttphnifpvislylmgevtndgsfriltpqgylrpvedv 360
 Oy 376 ATSDQDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARRIGFAVSACHVHDEFRTAAVEG 435
 |||
 Db 361 atsqddcykfaissqstgtvmgavimegfyvvdrrarkrigfavsachvhdefrtaaveg 420
 Oy 436 PFVTLDMEDCGYNIPOTDESTLMTIAYVMAAICATLPLCLMVCQWRCRLCRLRQHHDF 495
 |||
 Db 421 pfvtldmedcgyndipqtdestlmtiayvmaaaicafmlplclmvqcwrcrlcrlrqghddf 480
 Oy 496 ADDISLLK 503
 |||
 Db 481 addisllk 488

RESULT 4

AAB61334 4
 ID AAB61334 standard; Protein: 488 AA.

XX AAB61334;
 XX

DT 02-APR-2001 (first entry)
 XX

DE Memapsin 2 protein.
 XX

KW Memapsin 2; catalyst; Alzheimer's.
 XX

OS Homo sapiens.
 XX

PN WO200100663-A2.
 XX

PD 04-JAN-2001.
 XX

PF 27-JUN-2000; 2000WO-US17661.
 XX

PR 28-JUN-1999; 99US-0141363.
 PR 30-NOV-1999; 99US-0168060.
 PR 25-JAN-2000; 2000US-0177836.
 PR 27-JAN-2000; 2000US-0178368.
 PR 08-JUN-2000; 2000US-0210292.
 XX

XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.
 PA

XX Tang JUN, Lin X, Koelsch G;
 PI

XX WPI; 2001-102885/11.
 XX

XX Purified recombinant catalytically active memapsin 2, used to screen
 PT inhibitors of it, which are used to treat and prevent Alzheimer's
 PT disease -
 XX

PS Claim 2; Page 73-75; 86pp; English.
 XX

XX The present invention relates to a purified recombinant
 CC catalytically active memapsin 2. The invention may be used for
 CC isolating inhibitors which are used to treat or prevent
 CC Alzheimer's disease. The invention may also be used to screen
 CC for individuals more genetically prone to develop Alzheimer's
 CC disease.
 XX

QY 496 ADDISLLK 503
 DB 494 addisllk 501

RESULT 6
 AAB07896
 ID AAB07896 standard; Protein; 501 AA.
 XX AAB07896;
 AC AAB07896;
 DT 14-NOV-2000 (first entry)
 DE Amino acid sequence of a human beta-secretase enzyme.
 KW Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;
 KW amyloid plaque component; Alzheimer's disease; amyloidogenic disease;
 KW inhibitor.
 XX Homo sapiens.
 OS Homo sapiens.
 PN WO200047618-A2.
 XX 17-AUG-2000.
 PD 10-FEB-2000; 2000WO-US03819.
 PF 10-FEB-1999; 99US-0119571.
 PR 15-JUN-1999; 99US-0139172.
 XX (ELAN-) ELAN PHARM INC.
 PA Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;
 PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;
 XX WPI; 2000-533011/48.
 DR N-PSDB; AAAS9550, AAAS9551.
 XX Purified beta-secretase protein used in assays to discover inhibitors
 PT which can be used for the treatment of amyloidogenic diseases e.g.
 PT Alzheimer's disease -
 XX Claim 17; Fig 2A; 121pp; English.
 PS The specification describes a beta-secretase enzyme. The enzyme cleaves
 CC beta-amyloid precursor protein to produce beta-amyloid peptide. This
 CC enzyme is therefore implicated in the production of amyloid plaque
 CC components which accumulate in the brains of individuals afflicted with
 CC Alzheimer's disease. Inhibitors of beta-secretase are administered to
 CC a mammalian subject e.g. with Alzheimer's disease or Alzheimer's
 CC disease-like pathology to test if they maintain or improve cognitive
 CC ability or reduce the plaque burden. The compounds are used for the
 CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The
 CC present sequence represents a human beta-secretase enzyme.
 XX Sequence 501 AA:

Query Match 97.1%; Score 2587; DB 21; Length 501;
 Best Local Similarity 100.0%; Pred. No. 5.5e-255;
 Matches 488; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 AGVLPAGHQHGIPLRLSGLGAPLGLRLPRETDEEPEGRGSGFVMDNLRKSGQ 75
 DB 14 agvlpahgqhghlrlpslsglgaplglrlpretdeeepegrgrsgfvmvndnlrksqg 73

QY 76 GYVEMTGSPPQTILNLDVTGSSNFAGVGAAPHFLHRYTQRLSTYRDLRKGVVYPT 135
 DB 74 gyyvemtgsppqtinildvtgssnfavgaaphflhryyqrlsdyrdlrgvvypt 133

QY 136 QGKWEGLGTDLVSPHGNVTVRANIAITSDSKFFINGSNWEGILGLAYAEIARPDSS 195
 DB 134 qgkwegelgtdlvspghnvtvranaiaitesdkffingsnwegilglayaeiarpdds 193

QY 196 LEPRFSLVKOTHVPLNLSLQCGAGFPNOSSEVLASVGGSMIIGIDHSLVTSGLWTP 255
 DB 194 lepfslsvkothvplnlsfqlcgagfplnsgsevlavsggsmilgghslvtsglwtpt 253

QY 256 IRREWYEVIIVRVEINGQDLKMDCKEYNDKSIYDSGTTNLRPKKVFEEAAVKSIIKAS 315
 DB 254 irrewyeviivrveingqdlkmdckeyndksiydsgttnlrlpkvfeavksikaas 313

QY 316 STEKFPDGFNLCEQLVCHQAGTTPHNIFPVISLYLMGEVTNQSFRTITLPQOYLRPVEDV 375
 DB 314 stekfpdgfnlceqlvchqagttphnifpvlslylmgevtngsfritlpgqylrpvedv 373

QY 376 ATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARRKIGFAVSACHVHDEFRTAAVEG 435
 DB 374 atsqddcykfaissqstgtvmgavimegyvfvdrarkrigfavsachvhdefrtaaveg 433

QY 436 PFVTLDMEDCGYNIPQTDSTLMTIAYVMAAICALFMLPLCLMVCOMRCLRLRQHQHDDF 495
 DB 434 pfvtldmedcgyndipqtdstlmtiaayvmaaaicalfmlplclmvcqwrclrlrqghddf 493

QY 496 ADDISLLK 503
 DB 494 addisllk 501

RESULT 7
 AAY88425
 ID AAY88425 standard; Protein; 501 AA.
 AC AAY88425;
 DT 03-AUG-2000 (first entry)
 DE Human aspartyl protease 2 (a) (Asp2) amino acid sequence.
 KW Aspartyl protease; aspartase; amyloid precursor protein; APP; Asp 2;
 KW Alzheimer's disease; beta secretase site.
 OS Homo sapiens.
 XX WO200017369-A2.
 PN 30-MAR-2000.
 PD 23-SEP-1999; 99WO-US20881.
 PR 24-SEP-1998; 98US-0101594.
 XX (PHAA) PHARMACIA & UPJOHN CO.
 PI Gurney ME, Bienkowski MJ, Heinrikson RL, Parodi LA, Yan R;
 XX WPI; 2000-303209/26.
 DR N-PSDB; AAAL5662.
 XX New enzyme designated human aspartase useful in research into
 PT Alzheimer's Disease is capable of cleaving amyloid protein precursor at
 PT the beta secretase site to produce amyloid beta peptide -
 XX Claim 48; Fig 2; 183pp; English.
 XX This sequence represents the human aspartyl protease 2 (Asp2) amino acid
 CC sequence. The invention relates to a protease (e.g. Asp2) capable of
 CC cleaving the beta secretase site of amyloid precursor protein (APP). The
 CC protease contains a sequence encoding the amino acid sequence DTG and a
 CC sequence encoding DSG or DTG separated by 100-300 amino acids. When
 CC mutated the APP gene causes an autosomal dominant form of Alzheimer's
 CC disease. APP localises to the cell surface membrane and have a single
 CC C-terminal transmembrane domain. Proteolytic processing of APP produces
 CC the amyloid beta protein, which is possibly very important in Alzheimer's
 CC disease. The invention includes a nucleotide sequence encoding the

CC protease, a vector containing the nucleotide sequence, and a cell line
 CC comprising the vector. Methods for screening for inhibitors of beta
 CC secretase activity are also given in the invention. The human aspartase
 CC protein and nucleotide sequences and the methods for identifying
 CC inhibitors of the protease, are useful in the treatment of and research
 CC in to Alzheimer's disease.
 XX
 SQ Sequence 501 AA;

Query Match 97.0%; Score 2582; DB 21; Length 501;
 Best Local Similarity 99.8%; Pred. No. 1.8e-254;
 Matches 487; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 16 AGVLPAHQGTQHGIRLPLRSGLGGAPLGLRLPRETDEEPEEPGRGGSFVEMVDNLGRKSGQ 75
 DB 14 agvlpahgtqhgirplrsrglggaplgirlpretdeeepeepgrrgsfvemvdlrgrksqg 73
 QY 76 GYVEMTVGSPPTQLNLILVDTGSSNFVAGAAPHPFLHRYQRQLSSTYRDLRKGYVVPYT 135
 DB 74 gyyvemtvgspptqlnlilvdtgssnfavgaaphflhryyqrqlsstyrdlrkgyvvpvt 133
 QY 136 QGKWEGLGTDLVSPHGPNTVVRANIAITESDKFFINGSNWEGILGLAYAEIARPDSS 195
 DB 134 qgkweglgtdlvshpgnpvtvranaiaitesdkffingsnwegilglayaeiarpdds 193
 QY 196 LEPPFDSL VKOTHVNLFLSLQLCGAGFPLNQSEVLASVGGSMIIGIDHSLYTGSLWTP 255
 DB 194 lepffdslvkthvnlflslqcgagfplnqsevlasvggsmiigidhsltytgslywtp 253
 QY 256 IRREWYEVIIIVRVEINGODLKMCKEYNDKSIYDSGTTNLRPKKVFEEAAVKSIIKAAS 315
 DB 254 irrewyevliivrveingodlkmckeyndksiydsgttnlrlpkkvfeaaavksikaas 313
 QY 316 STEKPPDGFWLGEOLVCWQAGTTPWNIPFVISLYLMGEVTNQSFRTITLPQOYLRPVEDV 375
 DB 314 stekfpdgfwlgeqlvcwqagttwnipfvislylmgevtngsfrtitlpqqlrpvedv 373
 QY 376 ATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVDRAKRIGFAVSACHVHDEFRTAAVEG 435
 DB 374 atsqddcykfaisqsstgtvmgavimegyvvdrrarkrigfavsachvhdefrtaaveg 433
 QY 436 PFVTLDMEDCGNIPQTDSTLMTIAYVMAAICALFMLPLCLMVCQWCLRLCRLRQHDDF 495
 DB 434 pfvtldmedcgnipqtdstlmtiayvmaaaicalfmlplclmvcqwrclrlcrlrqhddf 493
 QY 496 ADDISLLK 503
 DB 494 addisllk 501

RESULT 8

AAW59807
 ID AAW59807 standard; Protein; 501 AA.
 XX
 AC AAW59807;

DT 26-OCT-1998 (first entry)
 XX

DE Amino acid sequence of human ASP2 (aspartic protease 2).

XX Human; ASP2; aspartic protease 2; agonist; antagonist; immunospecific;
 KW antibody; inhibition; Alzheimer's disease; cancer; proteinase;
 KW prohormone processing.

XX Homo sapiens.

XX EP855444-A2.

XX 29-JUL-1998.

XX 27-JAN-1998; 98EP-0300573.

XX

PR 28-JAN-1997; 97GB-0001684.
 XX (SMIK) SMITHKLINE BEECHAM CORP.
 PA (SMIK) SMITHKLINE BEECHAM PLC.
 XX Chapman CG, Murphy K, Powell DJ, Smith TS;
 PI WPI; 1998-389809/34.
 DR N-PSDB; AAV41696.

XX New nucleic acid encoding human aspartic protease 2 - used to treat,
 PT prevent and diagnose e.g. Alzheimer's disease, cancer and prohormone
 PT processing

XX Claim 1; Page 7; 26pp; English.

XX This is the amino acid sequence of the human ASP2 (aspartic protease
 CC family), used in the method of the invention. Agonists and
 CC antagonists for ASP2 immunospecific antibodies are used to treat
 CC conditions requiring increased or decreased activity or expression of
 CC ASP2 respectively. ASP2 is used to treat and diagnose e.g.
 CC Alzheimer's disease, cancer and prohormone processing and ASP2 or a
 CC fragment can be used to induce an immune response against the above
 CC conditions.

XX Sequence 501 AA;

Query Match 96.9%; Score 2581; DB 19; Length 501;
 Best Local Similarity 99.8%; Pred. No. 2.3e-254;
 Matches 487; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 16 AGVLPAHQGTQHGIRLPLRSGLGGAPLGLRLPRETDEEPEEPGRGGSFVEMVDNLGRKSGQ 75

DB 14 agvlpahgtqhgirplrsrglggaplgirlpretdeeepeepgrrgsfvemvdlrgrksqg 73

QY 76 GYVEMTVGSPPTQLNLILVDTGSSNFVAGAAPHPFLHRYQRQLSSTYRDLRKGYVVPYT 135

DB 74 gyyvemtvgspptqlnlilvdtgssnfavgaaphflhryyqrqlsstyrdlrkgyvvpvt 133

QY 136 QGKWEGLGTDLVSPHGPNTVVRANIAITESDKFFINGSNWEGILGLAYAEIARPDSS 195

DB 134 qgkweglgtdlvshpgnpvtvranaiaitesdkffingsnwegilglayaeiarpdds 193

QY 196 LEPPFDSL VKOTHVNLFLSLQLCGAGFPLNQSEVLASVGGSMIIGIDHSLYTGSLWTP 255

DB 194 lepffdslvkthvnlflslqcgagfplnqsevlasvggsmiigidhsltytgslywtp 253

QY 256 IRREWYEVIIIVRVEINGODLKMCKEYNDKSIYDSGTTNLRPKKVFEEAAVKSIIKAAS 315

DB 254 irrewyevliivrveingodlkmckeyndksiydsgttnlrlpkkvfeaaavksikaas 313

QY 316 STEKPPDGFWLGEOLVCWQAGTTPWNIPFVISLYLMGEVTNQSFRTITLPQOYLRPVEDV 375

DB 314 stekfpdgfwlgeqlvcwqagttwnipfvislylmgevtngsfrtitlpqqlrpvedv 373

QY 376 ATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVDRAKRIGFAVSACHVHDEFRTAAVEG 435

DB 374 atsqddcykfaisqsstgtvmgavimegyvvdrrarkrigfavsachvhdefrtaaveg 433

QY 436 PFVTLDMEDCGNIPQTDSTLMTIAYVMAAICALFMLPLCLMVCQWCLRLCRLRQHDDF 495

DB 434 pfvtldmedcgnipqtdstlmtiayvmaaaicalfmlplclmvcqwrclrlcrlrqhddf 493

QY 496 ADDISLLK 503

DB 494 addisllk 501

RESULT 9

AAW594769
 ID AAW594769 standard; Protein; 501 AA.
 XX

CC and as hybridization probes in diagnostic assays to test for the presence
 CC of beta-secretase DNA or RNA, such as in Alzheimer's disease, Down's
 CC syndrome, and amyloid angiopathy. The nucleotide sequence may also be
 CC used as anti-sense inhibitors of beta-secretase expression, in gene
 CC therapy of Alzheimer's disease, and for the identification of compounds
 CC that modulate beta-secretase activity. Antibodies to the beta-secretase
 CC protein may be used for in vitro and in vivo diagnostic purposes to
 CC detect the presence of beta-secretase polypeptide in a body fluid or cell
 CC sample. The present sequence represents the murine beta-secretase
 CC protein.

XX Sequence 501 AA;

Query Match 94.5%; Score 2517; DB 21; Length 501;
 Best Local Similarity 96.9%; Pred. No. 7.7e-248;
 Matches 473; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 16 AGVLPAGHTQHGIRLPLRSGLGAPLGLRLPRETDEEPEEGRGSGFVEMVDNLRGKSGQ 75
 DB 14 sgmlpaqgthgiriplrslgagplgrrlpretdeeseepgrgsfvmvndnlrgksqg 73
 QY 76 GYVEMTVGSPQTNLILVDTGSSNFAGAAPHPFLHRYQRLSTYRDLRKGVVYPT 135
 DB 74 gyyvemtvgspqtnilvdtgssnfavgaaphflhryyqrlslyrdlrgkgyvpyt 133
 QY 136 QGKWEGLGTLVSIPIHGNVTVRANIAAITESDKFFINGSNWEGILGLAYAEIARPD 195
 DB 134 qgkwegelgtlvsipghnvtvranaiaitesdkffingsnwegilglayaeiarpdds 193
 QY 196 LEPPFDSLKQTHVPLNLFSLQCGAGFPPLNQSEVLASVGGSMIIGDHSLYTGLWYTP 255
 DB 194 lepfdsllvkthipnlfslqlcgagfpplnqsealavsggsmiigdhsltyglwlytp 253
 QY 256 IRREWYEVIIIVRVEINGQDLKMDCKEYNDKSIYDSTNLRPKKVFEEAAVSIKAAS 315
 DB 254 irrewyevliivrveingdqlkmdckeyndksiydsgttnlrpkkvfeaaavsiikaas 313
 QY 316 STEKFPDGFGLGEQLVQWAGTTPWNIFPVISLYLMGEVTVNSFRITILPQQLRPVEDV 375
 DB 314 stekfpdgfwlgeqlvcwagttppwnifpvvislylmgevtngsfrtilpqqlrppedv 373
 QY 376 ATSDQDCYKFAISQSSTGTVMGAVIMEGYVYVDFRARKRIGFAVSACHVHDEFRTAAVEG 435
 DB 374 atsqdcdykvavqsstgtvmgavimegyvvyvdfdrarkrigfavsachvhdefrtaaveg 433
 QY 436 PFVTLDMEDCGYNIPQTDSTLMTIAYVMAATCALFMLPLCLMVQCWRCRLCRLQHQHDF 495
 DB 434 pfvtadmedcgyinipqtdstlmtiayvmaatacalfmplclmvqcwrcrlcrlrhqhddf 493
 QY 496 ADDISLLK 503
 DB 494 addisllk 501

RESULT 11
 ID AAY88427
 AC AAY88427; standard; Protein; 501 AA.
 AC AAY88427;
 XX

03-AUG-2000 (first entry)

XX Murine aspartyl protease 2 (a) (Asp2) amino acid sequence.
 DE Aspartyl protease; aspartase; amyloid precursor protein; APP; Asp 2;
 XX Alzheimer's disease; beta secretase site; mouse.
 KW Mus musculus.
 OS
 XX WO200017369-A2.
 PN
 XX 30-MAR-2000.
 PD

XX 23-SEP-1999; 99WO-US20881.
 XX 24-SEP-1998; 98US-0101594.
 PR (PHAA) PHARMACIA & UPJOHN CO.
 PA Gurney ME, Bienkowski MJ, Heinrikson RL, Parodi LA, Yan R;
 XX WPI; 2000-303209/26.
 DR N-PSDB; AAA15664.
 XX New enzyme designated human aspartase useful in research into
 PT Alzheimer's Disease is capable of cleaving amyloid protein precursor at
 PT the beta secretase site to produce amyloid beta peptide -
 PS Claim 105; Fig 4; 183pp; English.
 XX This sequence represents the murine aspartyl protease 2 (Asp2) amino acid
 CC sequence. The invention relates to a protease (e.g. Asp2) capable of
 CC cleaving the beta secretase site of amyloid precursor protein (APP). The
 CC protease contains a sequence encoding the amino acid sequence DTG and a
 CC sequence encoding DSG or DTG separated by 100-300 amino acids. When a
 CC mutated the APP gene causes an autosomal dominant form of Alzheimer's
 CC disease. APP localises to the cell surface membrane and have a single
 CC C-terminal transmembrane domain. Proteolytic processing of APP produces
 CC the amyloid beta protein, which is possibly very important in Alzheimer's
 CC disease. The invention includes a nucleotide sequence encoding the
 CC protease, a vector containing the nucleotide sequence, and a cell line
 CC comprising the vector. Methods for screening for inhibitors of beta
 CC secretase activity are also given in the invention. The human aspartase
 CC protein and nucleotide sequences and the methods for identifying
 CC inhibitors of the protease, are useful in the treatment of and research
 CC in to Alzheimer's disease.
 XX Sequence 501 AA;
 SQ

Query Match 94.5%; Score 2517; DB 21; Length 501;
 Best Local Similarity 96.9%; Pred. No. 7.7e-248;
 Matches 473; Conservative 6; Mismatches 9; Indels 0; Gaps 0;
 QY 16 AGVLPAGHTQHGIRLPLRSGLGAPLGLRLPRETDEEPEEGRGSGFVEMVDNLRGKSGQ 75
 DB 14 sgmlpaqgthgiriplrslgagplgrrlpretdeeseepgrgsfvmvndnlrgksqg 73
 QY 76 GYVEMTVGSPQTNLILVDTGSSNFAGAAPHPFLHRYQRLSTYRDLRKGVVYPT 135
 DB 74 gyyvemtvgspqtnilvdtgssnfavgaaphflhryyqrlslyrdlrgkgyvpyt 133
 QY 136 QGKWEGLGTLVSIPIHGNVTVRANIAAITESDKFFINGSNWEGILGLAYAEIARPD 195
 DB 134 qgkwegelgtlvsipghnvtvranaiaitesdkffingsnwegilglayaeiarpdds 193
 QY 196 LEPPFDSLKQTHVPLNLFSLQCGAGFPPLNQSEVLASVGGSMIIGDHSLYTGLWYTP 255
 DB 194 lepfdsllvkthipnlfslqlcgagfpplnqsealavsggsmiigdhsltyglwlytp 253
 QY 256 IRREWYEVIIIVRVEINGQDLKMDCKEYNDKSIYDSTNLRPKKVFEEAAVSIKAAS 315
 DB 254 irrewyevliivrveingdqlkmdckeyndksiydsgttnlrpkkvfeaaavsiikaas 313
 QY 316 STEKFPDGFGLGEQLVQWAGTTPWNIFPVISLYLMGEVTVNSFRITILPQQLRPVEDV 375
 DB 314 stekfpdgfwlgeqlvcwagttppwnifpvvislylmgevtngsfrtilpqqlrppedv 373
 QY 376 ATSDQDCYKFAISQSSTGTVMGAVIMEGYVYVDFRARKRIGFAVSACHVHDEFRTAAVEG 435
 DB 374 atsqdcdykvavqsstgtvmgavimegyvvyvdfdrarkrigfavsachvhdefrtaaveg 433
 QY 436 PFVTLDMEDCGYNIPQTDSTLMTIAYVMAATCALFMLPLCLMVQCWRCRLCRLQHQHDF 495
 DB 434 pfvtadmedcgyinipqtdstlmtiayvmaatacalfmplclmvqcwrcrlcrlrhqhddf 493

Qy 496 ADDISLLK 503
 Db 494 addisllk 501

RESULT 12

AA188426
 ID AAY88426 standard; Protein: 476 AA.

XX AC AAY88426;

DT 03-AUG-2000 (first entry)

XX Human aspartyl protease 2 (b) (Asp2) amino acid sequence.

DE DE Human aspartyl protease 2 (b) (Asp2) amino acid sequence.
 KW KW Aspartyl protease; aspartase; amyloid precursor protein; APP; Asp 2;
 KW KW Alzheimer's disease; beta secretase site.

XX OS Homo sapiens.

XX PN WO200017369-A2.

XX PD 30-MAR-2000.

XX PF 23-SEP-1999; 99WO-US20881.

XX PR 24-SEP-1998; 98US-0101594.

XX PA (PHAA) PHARMACIA & UPJOHN CO.

XX PI Gurney ME, Bienkowski MJ, Heinrichson RL, Parodi LA, Yan R;

XX DR WPI; 2000-303209/26.

XX DR N-PSDB; AAA15663.

PT New enzyme designated human aspartase useful in research into
 PT Alzheimer's disease is capable of cleaving amyloid protein precursor at
 PT the beta secretase site to produce amyloid beta peptide

XX PS Claim 51; Fig 3; 183pp; English.

XX This sequence represents the human aspartyl protease 2 (Asp2) amino acid
 CC sequence. The invention relates to a protease (e.g. Asp2) capable of
 CC cleaving the beta secretase site of amyloid precursor protein (APP). The
 CC protease contains a sequence encoding the amino acid sequence DTG and a
 CC sequence encoding DSG or DTG separated by 100-300 amino acids. When
 CC mutated the APP gene causes an autosomal dominant form of Alzheimer's
 CC disease. APP localises to the cell surface membrane and have a single
 CC C-terminal transmembrane domain. Proteolytic processing of APP produces
 CC the amyloid beta protein, which is possibly very important in Alzheimer's
 CC disease. The invention includes a nucleotide sequence encoding the
 CC protease, a vector containing the nucleotide sequence, and a cell line
 CC comprising the vector. Methods for screening for inhibitors of beta
 CC secretase activity are also given in the invention. The human aspartase
 CC protein and nucleotide sequences and the methods for identifying
 CC inhibitors of the protease, are useful in the treatment of and research
 CC in to Alzheimer's disease.

XX Sequence 476 AA;

Query Match 91.3%; Score 2432.5; DB 21; Length 476;
 Best Local Similarity 94.9%; Pred. No. 3e-239;
 Matches 463; Conservative 0; Mismatches 0; Indels 25; Gaps 1;

Qy 16 AGVLPAGTQHGIIRLPRLSRGLGAPLGLRLPRETDEEPEEPCRGSRFVEMVNLKSKSQ 75
 Db 14 agvlpahgtqhgiriprlsrslgagpirlpretdeepeepgrgsfvmvnlrgksq 73
 Qy 76 GYVEMTVGSPQTNILVDTGSSNFVACAPPHFLHRYIQRLSTYRDLKRGVVPVT 135
 Db 74 gyyvemtvgppqtnilvdtgssnfavgaapghflhryyqrlststydrlrkgyvpyt 133

Qy 136 QCKWEGELGTDLVSTPHGPNVTVRANIAAITESDKFFINGSNWEGILGLAYAEIARPPDS 195
 Db 134 qgkwegeigtldlvsiphgpnvtvranaiaaitesdkffingsnwegilglayaeiar---- 189
 Qy 196 LEPFDLSLVKQTHVPNLFSLQCGAGFPPLNQSEVLASVGGSMITGGIDHSLVTGSLWYTP 255
 Db 190 -----lcgagfplnqsevlasvggsmilggidhslvtgslwytp 228
 Qy 256 IRREWYEVIVRVEINGQDLKMDCKEYNDKSIYDSTTNLRLPKKPEAAVKSIKAAS 315
 Db 229 irrewyevilvrveingqdlkmdckeyndksiydsgttnlrlpkkvfeaaavksikaas 288
 Qy 316 STEKEPFDGFWLGEQLVCWQAGTTPWNIPVVISLYLMGEVNTOSFRITILPQOYLRPVEDV 375
 Db 289 stekfpdgdgfwlgeqlvcwqagtpwnipvislylmgevtngsfrtilpqyilrpvedv 348
 Qy 376 ATSQDCYKFAISQSSTGTVMGAVIMEGFYVVFDRKRIGFAVSACHVHDEFRTAAVGG 435
 Db 349 atsqddcykfaissgstgtvmgavimegyfvyvdrkrigfavsachvhdefrtaaveg 408
 Qy 436 PFVTLDMEDCGYNIPQTDTESTLMTIAYVMAAICALFMLPLCLMVQWRCRLRQOHDDF 495
 Db 409 pfvtldmedcgyinipqtdtestlmtiayvmaaaicalfmlplclmvqcwrcrlrqohddf 468
 Qy 496 ADDISLLK 503
 Db 469 addisllk 476

RESULT 13

AAAB07897
 ID AAB07897 standard; Protein: 456 AA.

XX AC AAB07897;

DT 14-NOV-2000 (first entry)

DE DE Active enzyme portion of human beta-secretase enzyme.

KW Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;
 KW amyloid plaque component; Alzheimer's disease; amyloidogenic disease;
 KW inhibitor; ss.

XX OS Homo sapiens.

XX PN WO2000047618-A2.

XX PD 17-AUG-2000.

XX PF 10-FEB-2000; 2000WO-US03819.

XX PR 10-FEB-1999; 99US-0119571.

XX PR 15-JUN-1999; 99US-0139172.

XX PA (ELAN-) ELAN PHARM INC.

XX PI Anderson JP, Basil G, Doane MT, Frigon N, John V, Power M;

XX PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;

XX DR WPI; 2000-533011/48.

XX Purified beta-secretase protein used in assays to discover inhibitors
 PT which can be used for the treatment of amyloidogenic diseases e.g.
 PT Alzheimer's disease

XX PS Claim 24; Fig 2B; 121pp; English.

XX The specification describes a beta-secretase enzyme. The enzyme cleaves
 CC beta-amyloid precursor protein to produce beta-amyloid peptide. This
 CC enzyme is therefore implicated in the production of amyloid plaque
 CC components which accumulate in the brains of individuals afflicted with
 CC Alzheimer's disease. Inhibitors of beta-secretase are administered to

CC a mammalian subject e.g. with Alzheimer's disease or Alzheimer's
CC disease-like pathology to test if they maintain or improve cognitive
CC ability or reduce the plaque burden. The compounds are used for the
CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The
CC present sequence represents the active enzyme portion of human
CC beta-secretase enzyme.
XX
SQ Sequence 456 AA;

Query Match 90.8%; Score 2419; DB 21; Length 456;
Best Local Similarity 100.0%; Pred. No. 6.6e-238;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 EYDEPEPEGRGSGFVEMVDNLGRSGGYVEMVNGSPQTLNLVDTGSSNFVAGAP 107
DB 1 EYDEPEPEGRGSGFVEMVDNLGRSGGYVEMVNGSPQTLNLVDTGSSNFVAGAP 60
QY 108 HPFLHRYQRLSSTYRDLRKGVVYPYTGQWEGELGTDLSIPHPNVTVRANIAAITE 167
DB 61 HPFLHRYQRLSSTYRDLRKGVVYPYTGQWEGELGTDLSIPHPNVTVRANIAAITE 120
QY 168 SKKFFINGSNWEGILGLAYAEIARPDLSLEPFDSLVKQTHVPNLFSLQCGAGFPLNQS 227
DB 121 SKKFFINGSNWEGILGLAYAEIARPDLSLEPFDSLVKQTHVPNLFSLQCGAGFPLNQS 180
QY 228 EYLASVGGSMIIIGGIDHSLYTGSLWYTPIRREWYVEIIVRVEINGQDLKMDCKEYNDK 287
DB 181 EYLASVGGSMIIIGGIDHSLYTGSLWYTPIRREWYVEIIVRVEINGQDLKMDCKEYNDK 240
QY 288 SIYDSTTNLRPLPKKVFEEAAVSIKAASTEKFPDGFVGLGEOLVCWQAGTTPWNTFPVIS 347
DB 241 SIYDSTTNLRPLPKKVFEEAAVSIKAASTEKFPDGFVGLGEOLVCWQAGTTPWNTFPVIS 300
QY 348 LYLMGEVNTQSPRITLPOQYLRPVEDVATSDODCKYKAISSNGTVMGAVIMESFYV 407
DB 301 LYLMGEVNTQSPRITLPOQYLRPVEDVATSDODCKYKAISSNGTVMGAVIMESFYV 360
QY 408 FDRARRKIGFAYSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDSTLMTIAYVMAAI 467
DB 361 FDRARRKIGFAYSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDSTLMTIAYVMAAI 420
QY 468 CALFMLPICLMVCQWRCLRCLQQHDDFADDISLLK 503
DB 421 CALFMLPICLMVCQWRCLRCLQQHDDFADDISLLK 456

RESULT 14

AY88431
ID AAY88431 standard; Protein; 446 AA.

XX AC AAY88431;

XX DT 03-AUG-2000 (first entry)

XX DE T7-caspase-human-pro-Asp-2(a)-deltaTM amino acid sequence.

XX KW Aspartyl protease; aspartase; amyloid precursor protein; APP; Asp 2;
KW Alzheimer's disease; beta secretase site;
KW T7-caspase-human-pro-Asp-2(a)-deltaTM.

XX OS Chimeric - Homo sapiens.

XX OS Chimeric - Bacteriophage t7.

XX PN W0200017369-A2.

XX PD 30-MAR-2000.

XX XX 23-SEP-1999; 99WO-US20881.

XX XX 24-SEP-1998; 98US-0101594.

XX XX (PHAA) PHARMACIA & UPJOHN CO.

XX Gurney ME, Bienkowski MJ, Heinrikson RL, Parodi LA, Yan R;
XX WPI; 2000-303209/26.
XX N-PSDB; AAA15668.
XX New enzyme designated human aspartase useful in research into
XX Alzheimer's Disease is capable of cleaving amyloid protein precursor at
XX the beta secretase site to produce amyloid beta peptide .
XX Example 9; Fig 6; 183pp; English.

PS This sequence represents a modified version of the human aspartase 2
XX (Asp2) amino acid sequence. The sequence is used in the bacterial
XX expression of human Asp2L. The invention relates to a protease
XX (e.g. Asp2) capable of cleaving the beta secretase site of amyloid
XX precursor protein (APP). The protease contains a sequence encoding the
XX amino acid sequence DTG and a sequence encoding DSG or DTG separated by
XX 100-300 amino acids. When mutated the APP gene causes an autosomal
XX dominant form of Alzheimer's disease. APP localises to the cell surface
XX membrane and have a single C-terminal transmembrane domain. Proteolytic
XX processing of APP produces the amyloid beta protein, which is possibly
XX very important in Alzheimer's disease. The invention includes a
XX nucleotide sequence encoding the protease, a vector containing the
XX nucleotide sequence, and a cell line comprising the vector. Methods for
XX screening for inhibitors of beta secretase activity are also given in the
XX invention. The human aspartase protein and nucleotide sequences and the
XX methods for identifying inhibitors of the protease, are useful in the
XX treatment of and research in to Alzheimer's disease.

XX SQ Sequence 446 AA;

Query Match 87.7%; Score 2336.5; DB 21; Length 446;
Best Local Similarity 97.8%; Pred. No. 1.7e-229;
Matches 445; Conservative 0; Mismatches 1; Indels 9; Gaps 1;

QY 1 MASMTGGQMGGRSGMAGVLPAGHTQHGIRLPLRSLGGLGAPLGLRPLRTDEPEEPGRRG 60

DB 1 masmtggqmggrgs-----tqhgirplrslgglgaplgirlpretdeepeegrgrg 51

QY 61 SFVEMVDNLGRCKSGGYVEMTVGSPQTLNLVDTGSSNFVAGAAPHPFHRYVQRLS 120

DB 52 sfvemvdnlrgksqgyvemtvgspqtlnlvdtgssnfavgaapfhlrhyvqrqls 111

QY 121 STYRDLRKGVVYPYTGQWEGELGTDLSIPHPNVTVRANIAAITEKFFINGSNWEG 180

DB 112 styrdlrkgvvyptytgqkwegeltldlvsiphgnvtvranaiaitesdkffingsnweg 171

QY 181 ILGLAYAEIARPDLSLEPFDSLVKQTHVPNLFSLQCGAGFPLNQSEVLASVGGSMIIIG 240

DB 172 ilglayaeiarpdlslepffdsllvkqthvpnlfsllhcgagfplnqsevlasvggsmiig 231

QY 241 GDHSLYTGSLWYTPIRREWYVEIIVRVEINGQDLKMDCKEYNDKSIYDSTTNLRPL 300

DB 232 gidhsltycslwytpirrewyveiiivveingqdlkmdckeyndksiydsgttnlrpl 291

QY 301 KKVFEAAVKSIIKAASTTEKFPDGFVGLGEOLVCWQAGTTPWNTFPVISLYLMGEVNTQSF 360

DB 292 kkvfeaaavksikaasstekfpdgfwlgeqlvcwqagttppwnifpvvislylmgevntqsf 351

QY 361 ITILPQOYLRPVEDVATSDODCKYKAISSNGTVMGAVIMESFYVDFDRARRKIGFAVS 420

DB 352 itilpqoylrpvedvatsqddcykfaisqstgtvmgavimegfvyvdfdrarkrigrfavs 411

QY 421 ACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQTD 455

DB 412 achvhdefrtaavegpfvltlmedcgyinipqtides 446

RESULT 15

AAY88432

ID AAY88432 standard; Protein; 459 AA.

XX AC AAY88432;
XX DT 03-AUG-2000 (first entry)
XX DE T7-caspase-human-pro-Asp-2(a)-deltaTM amino acid sequence.
XX KW Aspartyl protease; aspartase; amyloid precursor protein; APP; Asp 2;
XX KW Alzheimer's disease; beta secretase site;
XX KW T7-caspase-human-pro-Asp-2(a)-deltaTM.
XX OS Chimeric - Homo sapiens.
XX OS Chimeric - Bacteriophage t7.
XX PN WO200017369-A2.
XX PD 30-MAR-2000.
XX PF 23-SEP-1999; 99WO-US20881.
XX PR 24-SEP-1998; 98US-0101594.
XX PA (PHAA) PHARMACIA & UPJOHN CO.
XX PI Gurney ME, Bienkowski MJ, Heinrichson RL, Parodi LA, Yan R;
XX DR WPI; 2000-303209/26.
XX DR N-PSDB; AAA15669.
XX PT New enzyme designated human aspartase useful in research into
XX PT Alzheimer's disease is capable of cleaving amyloid protein precursor at
XX PT the beta secretase site to produce amyloid beta peptide -
XX PS Example 9; Fig 7; 183pp; English.
XX CC This sequence represents a modified version of the human aspartase 2
XX CC (Asp2) nucleotide sequence. The sequence is used in the bacterial
XX CC expression of human Asp2L. The invention relates to a protease
XX CC (e.g. Asp2) capable of cleaving the beta secretase site of amyloid
XX CC precursor protein (APP). The protease contains a sequence encoding the
XX CC amino acid sequence DTG and a sequence encoding DSG or DTG separated by
XX CC 100-300 amino acids. When mutated the APP gene causes an autosomal
XX CC dominant form of Alzheimer's disease. APP localises to the cell surface
XX CC membrane and have a single C-terminal transmembrane domain. Proteolytic
XX CC processing of APP produces the amyloid beta protein, which is possibly
XX CC very important in Alzheimer's disease. The invention includes a
XX CC nucleotide sequence encoding the protease, a vector containing the
XX CC screening for inhibitors of beta secretase activity are also given in the
XX CC invention. The human aspartase protein and nucleotide sequences and the
XX CC methods for identifying inhibitors of the protease, are useful in the
XX CC treatment of and research in to Alzheimer's disease.
SQ Sequence 459 AA;

Query Match 87.7%; Score 2335; DB 21; Length 459;
Best Local Similarity 97.4%; Pred. No. 2.5e-229;
Matches 447; Conservative 0; Mismatches 8; Indels 4; Gaps 2;

QY 1 MASWTGGQMGKGSMA-GVLPAH---GTQHGIRLPLRSLGGLGAPLGLRPRETDEEPEP 56
Db 1 masmtggqmgkrgsmtdspsregdgstqhgirplrslgglgagplglrlpretdeepeep 60
QY 57 GRGSGFVEMVNLKSGQYVVTGSPPTLNLVDTGSSNFAGCAAPHPFLHRYQ 116
Db 61 grgsgfvmvndlrsgsggyvvtgspptqnlvdtgssnfavgaapbflhryyq 120
QY 117 RQLSSTYRDLRKGVVYPYTGKWEGLGELTDLVSIHPGNVTVRANIAAITESDKFFINGS 176
Db 121 rqlsstyrdlrgkgyvpytgkwegelgtdlvsipgpnvtvrانياايتسدكففings 180
QY 177 NWEGLTGLAYAEIARPDSDSLEPFFDSLKQTHVPNLFSLQLCGAGFPLNQSEVLASVGS 236

Db 181 nwegilglayaeiarpddslpffdsllvkqthvpnlfsllhlcgagfplnqsevlaavgs 240
QY 237 MIIGIDHSLYTGSLWYTPIRREMYEYEVIVRVEINGODLKMCKEYNYDKSIVDSGTTN 296
Db 241 miigidhsllytgslywtpirrewyeyeviiirveingdldkmdckeynydkshivdsdgttn 300
QY 297 LRLPKKVFEEAAVKSIAKASSTKFPDGFGLGDLVCMQAGTTPWNIFPVVISLYLMGEVTN 356
Db 301 lrlpkkvfeaaavksikaasstekfpdgfwlgeqlvcwqagttcpwnifpvvislylmgevtc 360
QY 357 QSFRTILPQOYLRPVEDVATSDQDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIG 416
Db 361 qsfritilpqqylrpvedvatsqddcykfaisqsstgtvmgavimegfyvfvdrarkrig 420
QY 417 FAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPOTDES 455
Db 421 favsachvhdefrtaavegpfvtldmedcgyinipqt des 459

Search completed: September 6, 2001, 16:43:26
Job time: 355 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:39:33 ; Search time 113.12 Seconds
(without alignments)
91.557 Million cell updates/sec

Title: US-09-603-713-3

Perfect score: 2663

Sequence: 1 MASWTGGQMGGRGSMAGVLP.....CLRCRQOHDFADDISLLK 503

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

- 1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1175	44.1	518	3	US-08-999-723-2
2	1175	44.1	518	4	US-09-434-027-2
3	299.5	11.2	396	1	US-08-208-007A-13
4	299.5	11.2	396	4	US-09-032-523-9
5	296.5	11.1	412	1	US-08-208-007A-12
6	296.5	11.1	412	4	US-08-974-691-4
7	281	10.6	427	2	US-08-846-021A-8
8	279.5	10.5	458	6	5217891-15
9	273.5	10.3	409	1	US-08-360-673-6
10	270	10.1	410	1	US-08-088-633-2
11	270	10.1	410	1	US-08-245-756-2
12	270	10.1	410	1	US-08-441-750-2
13	270	10.1	410	2	US-08-441-751-2
14	270	10.1	410	5	PCT-US92-02521-2
15	250	9.4	349	4	US-09-032-523-3
16	232	8.7	398	1	US-08-328-314-2
17	232	8.7	398	1	US-08-731-045-2
18	216.5	8.1	397	3	US-09-079-415-2
19	215	8.1	419	4	US-08-974-691-3
20	203	7.6	427	1	US-07-958-222A-2
21	197	7.4	419	3	US-08-115-753-2
22	197	7.4	419	3	US-08-115-753-33
23	196.5	7.4	420	4	US-09-008-271A-4
24	196.5	7.4	420	4	US-08-974-691-8
25	196.5	7.4	430	1	US-08-535-237-2
26	194	7.3	395	1	US-08-723-938-3
27	194	7.3	395	2	US-09-080-538-3

28	194	7.3	445	4	US-08-974-691-6	Sequence 6, Appl
29	194	7.3	451	4	US-08-974-691-2	Sequence 2, Appl
30	193	7.2	330	3	US-08-115-753-1	Sequence 1, Appl
31	129.5	4.9	140	3	US-09-211-631-13	Sequence 13, Appl
32	129.5	4.9	140	4	US-09-265-628-13	Sequence 13, Appl
33	129.5	4.9	140	4	US-09-001-141-11	Sequence 11, Appl
34	97	3.6	1030	4	US-09-091-117-2	Sequence 2, Appl
35	95.5	3.6	280	4	US-09-160-246-14	Sequence 14, Appl
36	92.5	3.5	1097	2	US-08-680-326-39	Sequence 39, Appl
37	88	3.3	746	2	US-08-838-219B-6	Sequence 6, Appl
38	88	3.3	746	3	US-09-233-336A-6	Sequence 6, Appl
39	88	3.3	746	4	US-09-233-752A-6	Sequence 6, Appl
40	88	3.3	789	1	US-08-471-033-32	Sequence 32, Appl
41	88	3.3	789	2	US-08-471-044-32	Sequence 32, Appl
42	88	3.3	789	2	US-08-463-483A-32	Sequence 32, Appl
43	88	3.3	789	2	US-08-471-046A-32	Sequence 32, Appl
44	88	3.3	789	2	US-08-470-566B-32	Sequence 32, Appl
45	88	3.3	789	2	US-08-838-219B-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-999-723-2
; Sequence 2, Application US/08999723A
; Patent No. 6025180
; GENERAL INFORMATION:
; APPLICANT: Powell, David J.
; APPLICANT: Southan, Christopher
; APPLICANT: Chapman, Conrad G.
; APPLICANT: Evans, Joanne R.
; TITLE OF INVENTION: ASP1
; FILE REFERENCE: GH70262
; CURRENT APPLICATION NUMBER: US/08/999,723A
; CURRENT FILING DATE: 1997-10-06
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-999-723-2

Query Match 44.1%; Score 1175; DB 3; Length 518;
Best Local Similarity 47.1%; Pred. No. 2e-116;
Matches 228; Conservative 81; Mismatches 151; Indels 24; Gaps 5;
QY 20 PAHGTQHGRKLRSLGGLGAPLGLRLPRETDEPEEPGRGSGFVEMVDNLRGKSGGYV 79
DB 52 PAERHADGLALALEPALA-----SPAGAAFLAMVDNLQDGSGRGYL 94
QY 80 EMTVGPPOPTNLITVDTGSSNFAVGAAPHLHRYTORLSSTYRDLRKGVVYPYTOGW 139
DB 95 EMLIGTPQQLQILVDGTGSSNFAVAGTPHSYIDTFDTERSKGDFVTVKYGTSW 154
QY 140 EGEIGDLVSIPIHGPVNTVRANAAITESDKFFINGSNWEIGILGLAYAEIARPDSDLFP 199
DB 155 TGFVGEDLVTPKGFNTSFLVNIATIFESNFPLGIKKNIGILGLAYATLAKPSSLET 214
QY 200 FDSLVRKTHVNPVLFSLQCGAGFPLNOSVLA SVGSNITGGIDHSLYTSLWYTPTRR 259
DB 215 FDSLVTQANTPNVFSMOMCGALPVAGS---GTNGGSLVIGIEPSLYKGDINWYPIKEE 271
QY 260 WYVEIIVRVEINGQDLKMDCKEYNDKSI VDSCTTNLRPKKVFEEAVKSAASSTEK 319
DB 272 WYQIEILKLEIGQGSUNLDCREYNADKAL VDSGTTLLRLPQKVFDAVAVARSLIPE 331
QY 320 FPDGFVLGEOLVCWQAGCTTPWNIFPVISL YLMGVTNQSFRITILPOOYLRPVEDVATSQ 379
DB 332 FSDGFWTGSLACWTNSETPWSYFPKIS IYLRDENSRSFRITILPOLYIQPMWAGLNY 391

QY 380 DDCYKFAISOSSTGTVMGAVIMEGFVYVVEDRARKRIGCFVAVSACHVHDEFRTAAVEGPFVT 439
Db 392 -ECYREGISPTNALVIGTVMGAVIMEGFVYVVEDRARKRIGCFVAVSACHVHDEFRTAAVEGPFVT 450
QY 440 LDMEDCGYNIPOTDESTLMTIAYVMAAIC-ALFMLPLCLMVCMQWRCLRLRQOHDDFADD 498
Db 451 EDVASCNCPAQSLEPILMIVSYALMSVCGAILLVILVILLPFRQCR--RPRDPEVWVD 508
QY 499 ISLL 502
Db 509 ESSL 512

RESULT 2
US-09-434-427-2
; Sequence 2, Application US/09434427
; Patent No. 6162630
; GENERAL INFORMATION:
; APPLICANT: POWELL, DAVID J.
; APPLICANT: SOUTHAN, CHRISTOPHER
; APPLICANT: CHAPMAN, CONRAD G.
; APPLICANT: EVANS, JOANNE R.
; TITLE OF INVENTION: ASP1
; FILE REFERENCE: GH-70262-D1
; CURRENT APPLICATION NUMBER: US/09/434,427
; CURRENT FILING DATE: 1999-11-04
; EARLIER APPLICATION NUMBER: US 08/999,723
; EARLIER FILING DATE: 1997-10-06
; EARLIER APPLICATION NUMBER: UK 9626022.9
; EARLIER FILING DATE: 1996-12-14
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 518
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-09-434-427-2

Query Match 44.1%; Score 1175; DB 4; Length 518;
Best Local Similarity 47.1%; Pred. No. 2e-116;
Matches 228; Conservative 81; Mismatches 151; Indels 24; Gaps 5;

QY 20 PAHGTHQTRPLRSLGGLAPLGLPLRPTDEEPGRGRGSEVEMVMDNLKSGQGYV 79
Db 52 PAERHADGLALALEPALA-----SPAGAAFLAMVDNLQDGRGYVL 94
QY 80 EMTVGSPPOTLNLVDTGSSNFVAGAAHPFLHRYQRLSTYRDLRKGVVVPTQGW 139
Db 95 EMLGTPPKQLQLVDTGSSNFVAGATHSYIDTFDTERSTYRSKGFVDVTKYTGSM 154
QY 140 EGELGTDLVSIPIGPNVTVRANIAITESDKPFIINGSNWEGILGLAYAEIARPDLSLEPF 199
Db 155 TGFVGEDLVTPKGFNTSLVNIATIFESNFFLGKIKWNGILGLAYATLAKPSSLETF 214
QY 200 FDSLVKQTHVPLFSLQCGAGFPLNQSEVLASVGGSMIIGDHSLSYTGSLWYTPIRRE 259
Db 215 FDSLVQIANPVNFWMQMGAGLPVAGS---GTNGSGSLVGLGIEPSLXKGDIIWYTPIKEE 271
QY 260 WYVEVLIIVRVEINGODLKMCKEYNYDKSIVDSGTTNLRPKKVFEEAAVKSKAASSTEK 319
Db 272 WYQIETLLEIGGSLNDLREYNADRAIVDSGTTILRLPKQKVDADVAVARASLPIE 331
QY 320 PDGFWLGEQLVCMQAGTTPWNIIPVLSLYLMGEVYNGSFRITILPQOYLRPVEDVATSQ 379
Db 332 FSDGFWTGSQACLTNSETPNWSEYFPKISILRDENSSRSFRITILPQYIOPMPCAGLNY 391
QY 380 DDCYKFAISQSTGTVMGAVIMEGFVYVVEDRARKRIGCFVAVSACHVHDEFRTAAVEGPFVT 439
Db 392 -ECYREGISPTNALVIGTVMGAVIMEGFVYVVEDRARKRIGCFVAVSACHVHDEFRTAAVEGPFVT 450
QY 440 LDMEDCGYNIPOTDESTLMTIAYVMAAIC-ALFMLPLCLMVCMQWRCLRLRQOHDDFADD 498
Db 451 EDVASCNCPAQSLEPILMIVSYALMSVCGAILLVILVILLPFRQCR--RPRDPEVWVD 508

Db 451 EDVASCNCPAQSLEPILMIVSYALMSVCGAILLVILVILLPFRQCR--RPRDPEVWVD 508
QY 499 ISLL 502
Db 509 ESSL 512

RESULT 3
US-08-208-007A-13
; Sequence 13, Application US/08208007A
; Patent No. 5501969
; GENERAL INFORMATION:
; APPLICANT: HASTINGS, ET AL.
; TITLE OF INVENTION: Human Osteoclast-Derived Cathepsin
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/208,007A
; FILING DATE: March 8, 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: No. 5501969e
; FILING DATE: No. 5501969e
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-95
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
US-08-208-007A-13

Query Match 11.2%; Score 299.5; DB 1; Length 396;
Best Local Similarity 25.9%; Pred. No. 2.8e-23;
Matches 100; Conservative 65; Mismatches 148; Indels 73; Gaps 16;

QY 50 DEPEEPGRGRGSEVEMVMDNLKSGQGYVEMTVPQTLNLVDTGSSNFVAGAA---105
Db 63 DQSAKEP-----LNYLD-----MEYFGTISIGSPQNFVIFDTGSSNLWSPVSYCT 110
QY 106 APHPFLHRYQRLSTYRDLRKGVVVPTQGWELGTLDSYIPHPGNVTVRANIAI 165
Db 111 SPACKTHSRFQSPQSSTYSQPGSFSIQYGTGSLSGIIGADQSV-EGLTVVYVQOQFGESV 169
QY 166 TESDRFFINGSNWEGILGLAYAEIARPDLSLPPFDSLVKQTHVPLNLSLQCGAGFPLN 225
Db 170 TEPGQTFVD-AEFDGILGLGYPSLA--VGGVTPVDNMAQ-----NLVDLPMSFYMSN 222
QY 226 QSEVLASVGGSMIIGDHSLSYTGSLWYTPIRREYVEVLIIVRVEINGODLKMCKEYNY 285
Db 223 PE--GGAGSELIFGQYDHSFSGSLNWVPVTKQAYQWIALDNIQVG--TVMFCSE--G 275
QY 286 DKSIVDSGTTNLRPKKVFEEAAVKSKAASSTEKFPDGFGLGEQLVCMQAGTTPWNIIPV 345

Db 276 COAIYDVTSLTIGPSDKIKQLONAIGAAP-----VDGEYAVE-----CANLNVMPD 322
QY 346 ISLYLMGEVNTQSFRTILPOOYLRPVEDVATSDQCYKFAISQSTG----- 393
Db 323 VFTFTG-----VPYTLSPYAT--TLDFVDMQFC-----SSGFGOLDIHPPAGP 366
QY 394 -TVMGAVIMEGYVVDPRARKRIGFA 418
Db 367 LWILGDVFIRQFYSVDFRGNRRVGLA 392

RESULT 4
US-09-032-523-9
; Sequence 9, Application US/09032523
; Patent No. 6232454
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl
; APPLICANT: Baugh, Mariah
; TITLE OF INVENTION: HUMAN PROTEINASE MOLECULES
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/032.523
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0479 US
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 396 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Genbank
CLONE: 181994

US-09-032-523-9

Query Match 11.2%; Score 299.5; DB 4; Length 396;
Best Local Similarity 25.9%; Pred. No. 2.8e-23;
Matches 100; Conservative 65; Mismatches 148; Indels 73; Gaps 16;

QY 50 DEEPPEPGRGSGFVEMVDNLCKSGQGYVEMTVGSPQTLNILDVTGSSNFAVGA---- 105
Db 63 DQSAKEP-----LINYLD-----MEYFTISIGSPQNTVIFDTGSSNLWVPSVYCT 110
QY 106 AHPHFLHYRQQLSTYSRDLRGKVYVPTQGWKEGELGTLDSIPHPGNVTVRANIAI 165

Db 111 SPACKTHSRFPQSQSTYSQPCQSFISQYGTGSLSGIICADQSV-EGLTVVQQRGESV 169
QY 166 TESDKFFINGSNWEGILGLAYAEIARPDSDLEPFDFSLVKQTHVPNLFSLQLCAGGFPLN 225
Db 170 TEPGQTEVD-AEFDGLIGLGYPSLA--VGVTPVFDNMAQ-----NLVDLPMSFVYMSSN 222
QY 226 QSEVLASVGGSMIIIGIDHSILYTGSLWYTPIRREYYEVIIVRVEINGQDLKMDCKEYNY 285
Db 223 PE---GAGSELIFGQYDHSFSGSLNWPVTKQAYWQIALDNIQVGG--TVMFCSE--G 275
QY 286 DKSIVDSGTTNLRPKKVFEEAAVKSIKAASSTKPKDPDFWLGELQVCMQAGTTPWNIFPV 345
Db 276 COAIYDVTSLTIGPSDKIKQLONAIGAAP-----VDGEYAVE-----CANLNVMPD 322
QY 346 ISLYLMGEVNTQSFRTILPOOYLRPVEDVATSDQCYKFAISQSTG----- 393
Db 323 VFTFTG-----VPYTLSPYAT--TLDFVDMQFC-----SSGFGOLDIHPPAGP 366
QY 394 -TVMGAVIMEGYVVDPRARKRIGFA 418
Db 367 LWILGDVFIRQFYSVDFRGNRRVGLA 392

RESULT 5
US-08-208-007A-12
; Sequence 12, Application US/08208007A
; Patent No. 5501969
; GENERAL INFORMATION:
; APPLICANT: HASTINGS, ET AL.
; TITLE OF INVENTION: Human Osteoclast-Derived Cathepsin
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/208.007A
FILING DATE: March 8, 1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: No. 5501969e
FILING DATE: No. 5501969e
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-95
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 412 AMINO ACIDS
TYPE: AMINO ACID
STRANDEDNESS:
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
US-08-208-007A-12

Query Match 11.1%; Score 296.5; DB 1; Length 412;
Best Local Similarity 26.8%; Pred. No. 6.3e-23;
Matches 117; Conservative 66; Mismatches 163; Indels 91; Gaps 19;

QY 28 IRLPLR-----SGLGAPLGL-----RLPRETDEPEEPGRGSGFVEMVD 67

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 412 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
US-08-974-691-4

Query Match      11.1%; Score 296.5; DB 4; Length 412;
Best Local Similarity 26.8%; Pred. No. 6.3e-23;
Matches 117; Conservative 66; Mismatches 163; Indels 91; Gaps 19;

QY 22 VRIPLHKFTSIRRTMSEVGGSGVEDLIAKGPVSKYQAVPAVTE-----GPIPEVLK 72
DB 68 NLRGKSGGQYVEMTVGSPPTNLILVDTGSSNFAVGAAPHPFL-----HRYYQRLSS 121
DB 73 NYMDAQ---YYGEIGITPPQCFTVVFDTGSSNLWVPSHCKLLDIACWIHHKYNDSKSS 129
QY 122 TYRDLRKGVYVYPTQKWEGLGTLVSLP-----HGNVTVVRANIAIAITSDKFF 172
DB 130 TVVKNGTSDIHYGSGSLSGYLSQDTSVSPQSSASSALGCVKVERQVFGATKQPGIT 189
QY 173 INGSNWEGITGLAYAEIARPDSDLEPFDFSLVKQTHV-PNLFSLQLCGAGFPLNGSEVLA 231
DB 190 FTAAKFDGILGMAYPRIS--VNNVLVPFDNLMOQKLVQDNIFSYL-----SRDPDA 239
QY 232 SVGGSMIIGGDHSLYTGSLWTPTRRREYVEIIVRVEI-NGQDLKMDCKEYNDKSTIV 290
DB 240 QPGGELMLGGTDSKYKGSLSYLVNTRKAYWQVHLDQVEVASGLTL---CKE--GCEAIV 294
QY 291 DSGTTNLRPLPKKVFEEAAVKSIAKASTKFPDGFGLWGEOLV-CWQAGTTPWNIFPVISLY 349
DB 295 DTGTSMLWGPVDEVELQKAGAVPLIQ-----GEYMIPECKVST-----LPAITLK 341
QY 350 LMGEVTNQSFRITILPQOYLREPVEDVATSDDCYKFAISQ-----SSTGTVMGAVIMEGF 404
DB 342 LGG---KGYKLS--PEDYTLKVSQAGKTL--CLSGFMGMDIPPPSGPLWLILGDFVIGRY 393
QY 405 YVDFDRARKRIGFAYSA 421
DB 394 YTVFDRDNNRVGFAEAA 410

```

RESULT 6

```

US-08-974-691-4
; Sequence 4, Application US/08974691
; Patent No. 6225103
; GENERAL INFORMATION:
; APPLICANT: Keolsch, Gerald
; APPLICANT: Lin, Xindi
; APPLICANT: Tang, Jordan
; TITLE OF INVENTION: Cloning and Characterization of Napsin
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center, 1201 W. Peachtree
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/974,691
; FILING DATE: 20-NOV-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/031,196
; FILING DATE: 20-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/046,126
; FILING DATE: 09-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: OMFR 166
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-873-8794
; TELEFAX: 404-873-8795
; INFORMATION FOR SEQ ID NO: 4:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 412 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
US-08-974-691-4

Query Match      11.1%; Score 296.5; DB 4; Length 412;
Best Local Similarity 26.8%; Pred. No. 6.3e-23;
Matches 117; Conservative 66; Mismatches 163; Indels 91; Gaps 19;

QY 28 IRLPLR-----SGLGGAPLGL-----RLPRETDEEPEEGRGSGFVEMVD 67
DB 22 VRIPLHKFTSIRRTMSEVGGSGVEDLIAKGPVSKYQAVPAVTE-----GPIPEVLK 72
QY 68 NLRGKSGGQYVEMTVGSPPTNLILVDTGSSNFAVGAAPHPFL-----HRYYQRLSS 121
DB 73 NYMDAQ---YYGEIGITPPQCFTVVFDTGSSNLWVPSHCKLLDIACWIHHKYNDSKSS 129
QY 122 TYRDLRKGVYVYPTQKWEGLGTLVSLP-----HGNVTVVRANIAIAITSDKFF 172
DB 130 TVVKNGTSDIHYGSGSLSGYLSQDTSVSPQSSASSALGCVKVERQVFGATKQPGIT 189
QY 173 INGSNWEGITGLAYAEIARPDSDLEPFDFSLVKQTHV-PNLFSLQLCGAGFPLNGSEVLA 231
DB 190 FTAAKFDGILGMAYPRIS--VNNVLVPFDNLMOQKLVQDNIFSYL-----SRDPDA 239
QY 232 SVGGSMIIGGDHSLYTGSLWTPTRRREYVEIIVRVEI-NGQDLKMDCKEYNDKSTIV 290
DB 240 QPGGELMLGGTDSKYKGSLSYLVNTRKAYWQVHLDQVEVASGLTL---CKE--GCEAIV 294
QY 291 DSGTTNLRPLPKKVFEEAAVKSIAKASTKFPDGFGLWGEOLV-CWQAGTTPWNIFPVISLY 349
DB 295 DTGTSMLWGPVDEVELQKAGAVPLIQ-----GEYMIPECKVST-----LPAITLK 341
QY 350 LMGEVTNQSFRITILPQOYLREPVEDVATSDDCYKFAISQ-----SSTGTVMGAVIMEGF 404
DB 342 LGG---KGYKLS--PEDYTLKVSQAGKTL--CLSGFMGMDIPPPSGPLWLILGDFVIGRY 393
QY 405 YVDFDRARKRIGFAYSA 421
DB 394 YTVFDRDNNRVGFAEAA 410

```

RESULT 7

```

US-08-846-021A-8
; Sequence 8, Application US/08846021A
; Patent No. 5948682
; GENERAL INFORMATION:
; APPLICANT: Moloney, Maurice M.
; TITLE OF INVENTION: Preparation of Heterologous Proteins on
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BERESKIN & PARR
; STREET: 40 King Street West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5H 3Y2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,021A
; FILING DATE: April 25, 1997
; CLASSIFICATION: 800

```

```

Best Local Similarity 26.8%; Pred No. 4.9e-21;
Matches 95; Conservative 67; Mismatches 128; Indels 65; Gaps 15;

Qy 78 YVEMTVGSPQTLLNILDVTGSSNFVAAGHPFL-----HRYQRQLSSTYRDLRGV 130
      | : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 152 YFGIYLTPPQEFTVLDTGSSDFWV---PSYCKSNACKNHQRFDRKSTFQNLGKDL 208
Qy 131 VVPYTQKWEGELGTDLVSPHGPNVTVRANIAAITSDKFFINGSNWEGILGLAYAEIA 190
      | : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 209 SIHYGTSMOIGLDYDVTVSNIVDIOOTVLTGCEPDVF--TYAEPDGILGMAYPSLA 266
Qy 191 RPDSLSPFDLSLVKQTHV-PNFLSLQCGAGFPPLNQSEVLASVGGSMIIIGIDHSLYTG 249
      | : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 267 --SEYSIPVDENMNRHLVAQDLESVMYDRNG----QESMLT-----LCAIDPSYYTG 313
Qy 250 SLWYTPIRREWEYEVIIIVVEINGODLKMD--KREYNNDKSIVDSGTTNLRPKKVFEAA 307
      | : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 314 SLHWVPVTVOQYMQFTVDSVTISGVVACEGC-----QAILDGTSKLVGPSSDIINI 367
Qy 308 VK$TKAASSTEKFPDGFWLGP-QLVCHQAGTTPNNIFPVISLYLMGEVTNQSFRITILPQ 366
      | : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 368 QQAICATQNQ-----YGEFDIDCDNLSTMYPTWVF-----EINGKMYPLT--PS 408
Qy 367 QYLRPEVDVATSDDCY---KFAISQSSTGTMVGAVINMEGYVVVDFDRARKRIGFA 418
      | ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 409 AY-----TSQDQGECTSGFSQSENHSQGWILGDVFIREYYSVEDRANNLVGLA 455

RESULT          9
US-08-360-673-6
; Sequence 6, Application US/08360673
; Patent No. 5679544
; GENERAL INFORMATION:
; APPLICANT: Fleer, Reinhard
; APPLICANT: Fournier, Alain
; APPLICANT: Yeh, Patrice
; TITLE OF INVENTION: MODIFIED KLUYVEROMYCES YEASTS, THEIR
; TITLE OF INVENTION: PREPARATION AND USE
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rhone-Poulenc Rorer Inc.
; STREET: 500 Arcola Rd. 3C43
; CITY: Collegeville
; STATE: PA
; COUNTRY: USA
; ZIP: 19002
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/360.673
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/FR93/00623
; FILING DATE: 23-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/07785
; FILING DATE: 25-JUN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Julie K.
; REGISTRATION NUMBER: 38,619
; REFERENCE/DOCKET NUMBER: ST92040-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (610)454-3839
; TELEFAX: (610)454-3808
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 409 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

```


RESULT 12
US-08-441-750-2
; Sequence 2, Application US/08441750
; Patent No. 5691166
; GENERAL INFORMATION:
; APPLICANT: Gleeson, Martin A
; APPLICANT: Howard, Bradley D
; TITLE OF INVENTION: GENES WHICH INFLUENCE PICHIA PROTEOLYTIC
; TITLE OF INVENTION: ACTIVITY, AND USES THEREFOR
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/441,750
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/245,756
; FILING DATE: 16-May-1994
; CLASSIFICATION: 435

```

RESULT 13
US-08-441-751-2
; Sequence 2, Application US/08441751
; Patent No. 5831053
; GENERAL INFORMATION:
; APPLICANT: Gleeson, Martin A
; APPLICANT: Howard, Bradley D
; TITLE OF INVENTION: GENES WHICH INFLUENCE PICHA PROTEOLYTIC
; ACTIVITY, AND USES THEREFOR
; NUMBER OF INVENTION: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patenting Release #1.0
; Version #1.25

```


Job time: 123 sec

Search completed: September 6, 2001, 16:39:34

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:45:45 ; Search time 134.15 Seconds
(without alignments)
285.619 Million cell updates/sec

Title: US-09-603-713-3

Perfect score: 2863

Sequence: 1 MASWTGQGMGRGSMAGVLP.....CLRLRQHQHDFADDISLLK 503

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_58:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2587	97.1	501	2 A59090	aspartic proteinase
2	314	11.8	384	2 JC7574	pepsinogen A - Afr
3	308	11.6	385	3 JC7575	pepsinogen A - bul
4	306	11.5	387	2 B38302	pepsin (EC 3.4.23)
5	304.5	11.4	383	3 JC7573	pepsinogen C - Afr
6	304.5	11.4	388	1 S19682	pepsin A (EC 3.4.2)
7	302	11.3	384	2 A39314	gastricsin (EC 3.4)
8	301	11.3	382	1 PECH	pepsin A (EC 3.4.2)
9	299.5	11.2	396	2 A34401	cathepsin E (EC 3)
10	298.5	11.2	383	2 A41443	pepsin (EC 3.4.23)
11	296.5	11.1	412	1 KKHU	cathepsin D (EC 3)
12	296	11.1	387	2 C38302	pepsin (EC 3.4.23)
13	296	11.1	391	2 A43356	cathepsin E (EC 3)
14	295	11.1	387	2 D38302	pepsin (EC 3.4.23)
15	295	11.1	407	1 KHRD	cathepsin D (EC 3)
16	291.5	10.9	388	1 S19684	pepsin A (EC 3.4.2)
17	287.5	10.8	444	2 T24204	hypothetical prote
18	286	10.7	387	2 E38302	pepsin (EC 3.4.23)
19	285.5	10.7	380	2 I47176	chymosin (EC 3.4.2)
20	285.5	10.7	388	1 PEHU	pepsin A (EC 3.4.2)
21	285.5	10.7	388	1 PEKQAR	pepsin A (EC 3.4.2)
22	285.5	10.7	388	2 A30142	pepsin A (EC 3.4.2)
23	285.5	10.7	398	2 S66465	cathepsin E (EC 3)
24	285	10.7	389	2 JE0371	pepsin C (EC 3.4.2)
25	285	10.7	398	2 I51185	cathepsin D (EC 3)
26	284.5	10.7	410	1 KHMSD	cathepsin D (EC 3)
27	283.5	10.6	388	1 PEKQAJ	pepsin A (EC 3.4.2)
28	282.5	10.6	388	2 B30142	pepsin A (EC 3.4.2)
29	281.5	10.6	381	1 CM5HB	chymosin (EC 3.4.2)

pepsin A (EC 3.4.2)
pepsinogen A - com
cathepsin E (EC 3)
chymosin (EC 3.4.2)
gastricsin (EC 3.4)
pepsin (EC 3.4.23)
aspartic proteinase
cathepsin D (EC 3)
prochymosin - comm
candidapepsin (EC)
saccharopepsin (EC)
pepsinogen C - com
gastricsin (EC 3.4)
aspartic proteinase
gastricsin (EC 3.4)

30 278.5 10.5 386 1 PEPG
31 278 10.4 387 2 JC7245
32 277.5 10.4 396 2 S36865
33 273.5 10.3 381 1 CMBO
34 270.5 10.2 377 1 PEMQJ
35 270.5 10.2 389 2 A38302
36 270 10.1 376 2 I45856
37 268.5 10.1 344 1 KHPGD
38 268 10.1 396 2 T47207
39 267.5 10.0 381 2 JC7247
40 266 10.0 380 2 S03433
41 266 10.0 405 2 A5379
42 263.5 9.9 388 2 JC7246
43 263 9.9 394 2 B43356
44 261.5 9.8 387 2 A45117
45 261.5 9.8 388 2 A29937

RESULT 1
A59090
aspartic proteinase (EC 3.4.23.-) BACE precursor - human
N:Alternate names: beta-secretase; beta-site APP cleaving enzyme
C:Species: Homo sapiens (man)
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 11-May-2000
C:Accession: A59090
R:Vassar, R.; Bennett, B.D.; Babu-Khan, S.; Kahn, S.; Mendiaz, E.A.; Denis, P.; Teplio
M.A.; Biere, A.L.; Curran, E.; Burgess, T.; Louis, J.C.; Collins, F.; Treanor, J.; Ro
Science 286, 735-741, 1999
A:Title: beta-Secretase cleavage of Alzheimer's amyloid precursor protein by the tran
A:Reference number: A59090; MUID:20002972
A:Note: submitted to GenBank, September 1999
A:Accession: A59090
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-501 <VAS>
A:Cross-references: GB:AF190725; NID:g6118538; PIDN:AAF04142.1; PID:g6118539
C:Genetics:
A:Gene: BACE
C:Superfamily: beta-secretase
C:Keywords: Alzheimer's disease; aspartic proteinase; brain; glycoprotein; hydrolase;
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-45/Domain: propeptide #status predicted <PRO>
F:46-501/Product: acid proteinase BACE #status predicted <MAT>
F:461-477/Domain: transmembrane #status predicted <TRN>
F:93,289/Active site: Asp #status predicted
F:153,172,223,334/Binding site: carboxylate (Asn) (covalent) #status predicted
F:330-380/Disulfide bonds: #status predicted

Query Match 97.1%; Score 2587; DB 2; Length 501;
Best Local Similarity 100.0%; Pred. No. 6.5e+200;
Matches 488; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 AGVLPANGTQHGIRLPLRSLGGLGAPLGLRLPRETDEPEPEGRGRGSFVENVDNLRKSGQ 75
Db 14 AGVLPANGTQHGIRLPLRSLGGLGAPLGLRLPRETDEPEPEGRGRGSFVENVDNLRKSGQ 73

QY 76 GYVEMTVGSPPTNLILVDYTGSSNFVAGAAPHPFLHRYYQRLSTSYRDLRGVYVPT 135
Db 74 GYVEMTVGSPPTNLILVDYTGSSNFVAGAAPHPFLHRYYQRLSTSYRDLRGVYVPT 133

QY 136 QKWEGELGTDLSIPHPGNTVTRANIAITESDKFFINGSNWEGILGLAYAIARPDDS 195
Db 134 QKWEGELGTDLSIPHPGNTVTRANIAITESDKFFINGSNWEGILGLAYAIARPDDS 193

QY 196 LEPPFDSLVKQTHVPLNLSLQICGAGFPLNQSVLASVGGSMIIGGIDHSLYTGSLWYTP 255
Db 194 LEPPFDSLVKQTHVPLNLSLQICGAGFPLNQSVLASVGGSMIIGGIDHSLYTGSLWYTP 253

QY 256 IRREWTYEVIIIRVEINGQDLKMDCKEYNDKSIDVDSGTTNLRPKKVFEEAVKSKAAS 315

```
|||||
Db 254 IRREWEYEVIIVRVEINGDLMCKCKENYDKSVDSGTTNLRPKKVFEEAAVSIKAAS 313
QY 316 STEKFPDGFWMGEQLVCWQAGTTPNIFPVISLYLMGEVTNOSFRITILPOQYLRPVEDV 375
|||||
Db 314 STEKFPDGFWMGEQLVCWQAGTTPNIFPVISLYLMGEVTNOSFRITILPOQYLRPVEDV 373
QY 376 AYSQDDCKYFAISQSSTGTVMGAVTMEGFYVDFRARRKRGFAVSACHVHDEFRTAAVEG 435
|||||
Db 374 AYSQDDCKYFAISQSSTGTVMGAVTMEGFYVDFRARRKRGFAVSACHVHDEFRTAAVEG 433
QY 436 PFVTLDMEDCGYNIPQTDSTLMTAYMAAICALFMLPLCLMVQCWRCLRLRQHQHDF 495
|||||
Db 434 PFVTLDMEDCGYNIPQTDSTLMTAYMAAICALFMLPLCLMVQCWRCLRLRQHQHDF 493
QY 496 ADDISLLK 503
|||||
Db 494 ADDISLLK 501

RESULT 2
JC7574
pepsinogen A - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C>Date: 31-Mar-2001 #sequence_revision 31-Mar-2001 #text_change 31-Mar-2001
C:Accession: JC7574; PC7119
R:Ikuzawa, M.; Inokuchi, T.; Kobayashi, K.; Yasumasu, S.
J. Biochem. 129, 147-153, 2001
A:Title: Amphibian pepsinogens: Purification and characterization of Xenopus pepsinogens
A:Reference number: JC7573; PMID:11134969
A:Contents: Stomach
A:Accession: JC7574
A:Molecule type: mRNA
A:Residues: 1-384 <IKU>
A:Cross-references: DDBJ:AB045380
A:Accession: PC7119
A:Molecule type: protein
A:Residues: 16-35; 57-76 <IK2>
C:Comment: This protein is a zymogen for gastric aspartic proteinase, with pepsin-like
C:Genetics:
A:Gene: Pga
C:Keywords: stomach; zymogen

Query Match 11.8%; Score 314; DB 3; Length 384;
Best Local Similarity 25.1%; Pred. No. 1.8e-17;
Matches 107; Conservative 69; Mismatches 154; Indels 96; Gaps 18;

QY 28 IRLPLRSLGGAPLGLRLPRETDEPEEPGRGSGFVE-----HVDNLR 70
|||||
Db 17 VKVPLRKG-----ESFRNRPRQLGLLDYLKKNYPNASKYFPTLAQSSAETIQ 65

QY 71 GKSGQGYVEMTGPSPQTLNLIIVDTGSSNFAVGAAPHPL-----HRYQRLSSY 123
|||||
Db 66 NYMDIEYGTISGTTPQETVIFDGSANLWV---PSYCSQAQCSNHNRFNPQSSIF 122

QY 124 RDLRKGYVYPYQCKWEGELGTLVSIHPGNVTVRANIAITESDK--FFINGSNWEGIL 182
|||||
Db 123 QATNTPTVSIQYGTGSMGFLGYDTLQV---GNIQISNQMFGLSEPGSFLYSPFDGIL 179

QY 183 GLAYAEIARDDSLRPFDFSLVKQTHVP-NLFSLQLCGAGFPLNQSEVLASVCGSMILIG 241
|||||
Db 180 GLAPPSTA--SSQATPVDFNMWSQGLIPQLNLFSLSSDG-----QTGSYVLFEG 227

QY 242 IDLSLYTGSLSWYPIREWEYEVIIVRVEINGQDL--KMDCKEYNDKSVDSGTTNLR 299
|||||
Db 228 VDSNYFSGSLNWPLTAETWTITLDSVLSINGQVIACSSQ-----QAIVDTGTSIMTG 281

QY 300 PKKVFEAAVSIKAASTEKFPDGMWGEQLV-CWQAGTTPNIFPVISLYLMGEVTNOS 358
|||||
Db 282 PSTPI-ANIQNYIGASQDSN-----GQYVINCNNISNMPTIVF----- 318

QY 359 FRITILPOQY-LRPVEDVATSQDDCYK-FAISQSSGTG-----VMGAVIMBGFYVDFRARR 412
```

```
|||||
Db 319 ---TINGVQYPLSFSAVYRQKQCSGSGFQAMNLPNTSGDLWILGDFIROYFTVDFRAN 375
QY 413 KRIGFA 418
|||||
Db 376 NYVAIA 381

RESULT 3
JC7575
pepsinogen A - bullfrog
C:Species: Rana catesbeiana (bullfrog)
C>Date: 31-Mar-2001 #sequence_revision 31-Mar-2001 #text_change 31-Mar-2001
C:Accession: JC7575
R:Ikuzawa, M.; Inokuchi, T.; Kobayashi, K.; Yasumasu, S.
J. Biochem. 129, 147-153, 2001
A:Title: Amphibian pepsinogens: Purification and characterization of Xenopus pepsinogen
A:Reference number: JC7573; PMID:11134969
A:Contents: Stomach
A:Accession: JC7575
A:Molecule type: mRNA
A:Residues: 1-385 <IKU>
A:Cross-references: DDBJ:AB045376
C:Comment: This protein is a zymogen for gastric aspartic proteinase, with pepsin-like
C:Genetics:
A:Gene: Pga
C:Keywords: stomach; zymogen

Query Match 11.6%; Score 308; DB 3; Length 385;
Best Local Similarity 27.8%; Pred. No. 5.4e-17;
Matches 99; Conservative 64; Mismatches 133; Indels 60; Gaps 15;

QY 77 YVEMTVGSPQTLNLIIVDTGSSNFAVG----AAPHPFLHRYQRLSSTYRDLRKGVV 132
|||||
Db 73 YFGTISGTTPQSFVIFDTGSSNLWVPSVYCSQACTNHHMFNQSSSTQATNTPTVSI 132

QY 133 PYTQCKWEGELGTLVSIHPGNVTVRANIAITESDK--FFINGSNWEGILGLAYAEIAR 191
|||||
Db 133 QYGTGSMGFLGYDTVQV---GNIQITNQIFGLSQSEPSFLYSPFDGILGLAPPSLA- 188

QY 192 PDSLEPEFDSLVKQTHVP-NLFSLQLCGAGFPLNQSEVLASVCGSMIIGDHLSTGTS 250
|||||
Db 189 -SSQATPVDFNMWNOGLIPQDLFSVLSQSG-----QSGFVLEGGVDTSYTGN 237

QY 251 LMWYPIREWEYEVIIVRVEINGQDLKM--DCKEYNDKSVDSGTTNLRPKKVFEEAV 308
|||||
Db 238 LNWVPLTAETWTITVDSISIGQVIACSSG-----SAIVDTGTSIAGP-----STPI 287

QY 309 KSIKAASSTEKFPDGMWGEQLVCKQAGTTPNIFPVISLYLMGEVTNOSFRITILPOQY 368
|||||
Db 288 ANIQYIGANQDSNGQYV---INCNNISNMPTVVF-----TINGVQY 326

QY 369 LRPVED-VATSQDDC---YKFAISQSSGTG---VMGAVIMEGFYVDFRARRKIGFA 418
|||||
Db 327 PLPASAYVRQSQSCTSGFQAMNLPNTSGDLWILGDFIREYVYVDFRANNVAMA 382

RESULT 4
B38302
pepsin (EC 3.4.23.-) II-1 precursor - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C>Date: 14-Jun-1991 #sequence_revision 20-Sep-1991 #text_change 23-Feb-1997
C:Accession: B38302
R:Kageyama, T.; Tanabe, K.; Koizumi, O.
J. Biol. Chem. 265, 17031-17038, 1990
A:Title: Structure and development of rabbit pepsinogens. Stage-specific zymogens, nu
A:Reference number: A38302; MUID:91009127
A:Accession: B38302
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-387 <KAG>
A:Cross-references: GB:M59235; GB:J05638
```

Db 75 TPPQNFVLFDTGSSNLWVASTYCQSQACTNHPL---FNPSQSSSTYSSNQQFSLQYGT 130

1. The first part of the document is a list of names and their corresponding page numbers. The names are listed in a single column, and the page numbers are listed in a single column to the right of the names. The names are: J. A. B. C. D. E. F. G. H. I. J. K. L. M. N. O. P. Q. R. S. T. U. V. W. X. Y. Z. The page numbers are: 1. 2. 3. 4. 5. 6. 7. 8. 9. 10. 11. 12. 13. 14. 15. 16. 17. 18. 19. 20. 21. 22. 23. 24. 25. 26. 27. 28. 29. 30. 31. 32. 33. 34. 35. 36. 37. 38. 39. 40. 41. 42. 43. 44. 45. 46. 47. 48. 49. 50. 51. 52. 53. 54. 55. 56. 57. 58. 59. 60. 61. 62. 63. 64. 65. 66. 67. 68. 69. 70. 71. 72. 73. 74. 75. 76. 77. 78. 79. 80. 81. 82. 83. 84. 85. 86. 87. 88. 89. 90. 91. 92. 93. 94. 95. 96. 97. 98. 99. 100. 101. 102. 103. 104. 105. 106. 107. 108. 109. 110. 111. 112. 113. 114. 115. 116. 117. 118. 119. 120. 121. 122. 123. 124. 125. 126. 127. 128. 129. 130. 131. 132. 133. 134. 135. 136. 137. 138. 139. 140. 141. 142. 143. 144. 145. 146. 147. 148. 149. 150. 151. 152. 153. 154. 155. 156. 157. 158. 159. 160. 161. 162. 163. 164. 165. 166. 167. 168. 169. 170. 171. 172. 173. 174. 175. 176. 177. 178. 179. 180. 181. 182. 183. 184. 185. 186. 187. 188. 189. 190. 191. 192. 193. 194. 195. 196. 197. 198. 199. 200. 201. 202. 203. 204. 205. 206. 207. 208. 209. 210. 211. 212. 213. 214. 215. 216. 217. 218. 219. 220. 221. 222. 223. 224. 225. 226. 227. 228. 229. 230. 231. 232. 233. 234. 235. 236. 237. 238. 239. 240. 241. 242. 243. 244. 245. 246. 247. 248. 249. 250. 251. 252. 253. 254. 255. 256. 257. 258. 259. 260. 261. 262. 263. 264. 265. 266. 267. 268. 269. 270. 271. 272. 273. 274. 275. 276. 277. 278. 279. 280. 281. 282. 283. 284. 285. 286. 287. 288. 289. 290. 291. 292. 293. 294. 295. 296. 297. 298. 299. 300. 301. 302. 303. 304. 305. 306. 307. 308. 309. 310. 311. 312. 313. 314. 315. 316. 317. 318. 319. 320. 321. 322. 323. 324. 325. 326. 327. 328. 329. 330. 331. 332. 333. 334. 335. 336. 337. 338. 339. 340. 341. 342. 343. 344. 345. 346. 347. 348. 349. 350. 351. 352. 353. 354. 355. 356. 357. 358. 359. 360. 361. 362. 363. 364. 365. 366. 367. 368. 369. 370. 371. 372. 373. 374. 375. 376. 377. 378. 379. 380. 381. 382. 383. 384. 385. 386. 387. 388. 389. 390. 391. 392. 393. 394. 395. 396. 397. 398. 399. 400. 401. 402. 403. 404. 405. 406. 407. 408. 409. 410. 411. 412. 413. 414. 415. 416. 417. 418. 419. 420. 421. 422. 423. 424. 425. 426. 427. 428. 429. 430. 431. 432. 433. 434. 435. 436. 437. 438. 439. 440. 441. 442. 443. 444. 445. 446. 447. 448. 449. 450. 451. 452. 453. 454. 455. 456. 457. 458. 459. 460. 461. 462. 463. 464. 465. 466. 467. 468. 469. 470. 471. 472. 473. 474. 475. 476. 477. 478. 479. 480. 481. 482. 483. 484. 485. 486. 487. 488. 489. 490. 491. 492. 493. 494. 495. 496. 497. 498. 499. 500. 501. 502. 503. 504. 505. 506. 507. 508. 509. 510. 511. 512. 513. 514. 515. 516. 517. 518. 519. 520. 521. 522. 523. 524. 525. 526. 527. 528. 529. 530. 531. 532. 533. 534. 535. 536. 537. 538. 539. 540. 541. 542. 543. 544. 545. 546. 547. 548. 549. 550. 551. 552. 553. 554. 555. 556. 557. 558. 559. 560. 561. 562. 563. 564. 565. 566. 567. 568. 569. 570. 571. 572. 573. 574. 575. 576. 577. 578. 579. 580. 581. 582. 583. 584. 585. 586. 587. 588. 589. 590. 591. 592. 593. 594. 595. 596. 597. 598. 599. 600. 601. 602. 603. 604. 605. 606. 607. 608. 609. 610. 611. 612. 613. 614. 615. 616. 617. 618. 619. 620. 621. 622. 623. 624. 625. 626. 627. 628. 629. 630. 631. 632. 633. 634. 635. 636. 637. 638. 639. 640. 641. 642. 643. 644. 645. 646. 647. 648. 649. 650. 651. 652. 653. 654. 655. 656. 657. 658. 659. 660. 661. 662. 663. 664. 665. 666. 667. 668. 669. 670. 671. 672. 673. 674. 675. 676. 677. 678. 679. 680. 681. 682. 683. 684. 685. 686. 687. 688. 689. 690. 691. 692. 693. 694. 695. 696. 697. 698. 699. 700. 701. 702. 703. 704. 705. 706. 707. 708. 709. 710. 711. 712. 713. 714. 715. 716. 717. 718. 719. 720. 721. 722. 723. 724. 725. 726. 727. 728. 729. 730. 731. 732. 733. 734. 735. 736. 737. 738. 739. 740. 741. 742. 743. 744. 745. 746. 747. 748. 749. 750. 751. 752. 753. 754. 755. 756. 757. 758. 759. 760. 761. 762. 763. 764. 765. 766. 767. 768. 769. 770. 771. 772. 773. 774. 775. 776. 777. 778. 779. 780. 781. 782. 783. 784. 785. 786. 787. 788. 789. 790. 791. 792. 793. 794. 795. 796. 797. 798. 799. 800. 801. 802. 803. 804. 805. 806. 807. 808. 809. 810. 811. 812. 813. 814. 815. 816. 817. 818. 819

7

QY 217 LCGAGFLNQSEVLASVSGSMIIGDHSLSLTGSLWYTPIRREYVEVLIIVRVEINGDQL 276
DB 217 LSAD-----DQS-----GSAVIFGIDSYTSLWVPSVEGYWQISVDSITMNGKTI 266
QY 277 --KMDCKEYNDKSIDVSTLRLPKVKVFAAASIKAAASSTEFKPDGFWLGEOLV-CW 333
DB 267 ACAKGC-----QAIIVDTGSLTGTSPIANIQSDIGASENSD-----GENVVS 312
QY 334 QAGTTPWNIFFVISLYLMGEVNTNQSFRTILPQY-LRPVEDVATSDQCYK-----FAI 387
DB 313 AIISLPDIVE-----TINGVQYPLPPSAVILQSGSGTSGFGMDVP 354
QY 388 SOSSTGVMGAVIMEGFYVVPDRKRIGFA 418
DB 355 TESGELWILGDVFIROYFTVDFRANNOVGLA 385

RESULT 7
A39314
C:Species: Rana catesbeiana (bullfrog)
C:Date: 19-Jun-1992 #sequence_revision 19-Jun-1992 #text_change 22-Jun-1999
C:Accession: A39314
R:Yakabe, E.; Tanji, M.; Ichinose, M.; Goto, S.; Mikii, K.; Kurokawa, K.; Ito, H.; Kageya
J. Biol. Chem. 266, 22436-22443, 1991
A:Title: Purification, characterization, and amino acid sequences of pepsinogens and pep
A:Reference number: A39314; MUID:92042186
A:Accession: A39314
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-384 <YAK>
A:Cross-references: GB:M73750; NID:g213687; PIDN:AAA49530.1; PID:g213688
C:Superfamily: pepsin
C:Keywords: aspartic proteinase; hydrolase; protein digestion

Query Match 11.3%; Score 302; DB 2; Length 384;
Best Local Similarity 24.5%; Pred. No. 1.6e-16;
Matches 105; Conservative 65; Mismatches 147; Indels 112; Gaps 17;

QY 26 HGIRLPLRSLGGAPLGLRLPRETDEPEPGR--GSFVEMVDNLRGSGQGYVEMIV 83
DB 35 HGIKAPY-----VDPATKYNNFNATAFEPLANYMDMSYGEISI 73
QY 84 GSPQTLNILDVGTSSNFAVGAAPHPFL-----HRYQRLSSYRDLRKGVVPTQ 136
DB 74 GTPPNFLVLVFDGSSNLW---PSYQSQOACTNHPQFNPSOSSSYSSNQOQFSLOYGT 130
QY 137 GKWEGELGTLVSIPIHGPNTVVRANIA-----AITESDKFFINGSNWEGILGLAYAE 188
DB 131 GSLTGILGYDTVQI-----QNTAISQOEGLSVTEPGTNFVY-AQFDGILGLAYPS 180
QY 189 IARPDLSLEPFDFSLVKQTHVN-LPFLQICGAGFPLNQSEVLASVSGSMIIGDHSLY 247
DB 181 IA--EGGATTVMQMIQNLINQPLFAYLSGOONSQ-----GGEVAFGVDQNY 230
QY 248 TGLSWYTPIRREYVEVLIIVRVEINGD---LMDCKEYNDKSIDVSGTTLRLPKKVF 304
DB 231 SGQIYTPVTSETWQIGQFVNGQATWCSCQCC-----QGIVDTGSLTATQSVF 284
QY 305 EAAVKSIAASSTEFKPDGFWLGEOLV-CWQAGTTPWNIFFVI-----SLYLMGEVT 355
DB 285 SSLMQSIGAQDQN-----GGYAVSCNIOQLPTISFTISGVSPPLPPSAVILQONS 336
QY 356 NQ---SFRITILPQYLRPVEDVATSDQCYKFAISQSTGTCTVMGAVIMEGFYVVPDRAR 412
DB 337 GYCTIGIMPTLYPLSQNGOPL-----WILGDVFLRQYTSVIDLGN 375
QY 413 KRIGFAVA 421
DB 376 NOVGFANAA 384

RESULT 8
PECH
Pepsin A (EC 3.4.23.1) precursor - chicken
N:Alternate names: pepsinogen A
C:Species: Gallus gallus (chicken)
C:Date: 18-Apr-1984 #sequence_revision 01-Dec-2000 #text_change 01-Dec-2000
C:Accession: JE0370; A00984
R:Sakamoto, N.; Saiga, H.; Yasugi, S.
Biochem. Biophys. Res. Commun. 250, 420-424, 1998
A:Title: Analysis of temporal expression pattern and cis-regulatory sequences of chic
A:Reference number: JE0370; MUID:98440813
A:Accession: JE0370
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-382 <SAK>
A:Cross-references: GB:AB025281; NID:g4589837; PIDN:BAA76891.1; PID:g4589838
R:Baudys, M.; Kostka, V.
Eur. J. Biochem. 136, 89-99, 1983
A:Title: Covalent structure of chicken pepsinogen.
A:Reference number: A00984; MUID:84004412
A:Accession: A00984
A:Molecule type: protein
A:Residues: 16-87, S', 89-382 <BAU>
C:Superfamily: pepsin
C:Keywords: aspartic proteinase; gastric juice; glycoprotein; hydrolase; protein dige
F:16-57/Domain: activation peptide #status experimental <APT>
F:58-382/Product: pepsin A #status predicted <MAT>
F:92-275/Active site: Asp #status predicted
F:105-110,266-270,305-338/Disulfide bonds: #status experimental
F:128/Binding site: carbohydrate (Asn) #status experimental

Query Match 11.3%; Score 301; DB 1; Length 382;
Best Local Similarity 24.0%; Pred. No. 2e-16;
Matches 88; Conservative 68; Mismatches 126; Indels 84; Gaps 13;

QY 77 YIVEMTVGSPQTLNILDVGTSSNFAVGAAPHPFL-----HRYQRLSSYRDLRK 129
DB 74 YXGTISIGTPQDFTVIFDTGSSNLW---PSYCKSSACSNHKKRDPKSKSTVYSTNET 130
QY 130 VVVPYTGKWECELTGLVSIPIHGPNTVVRANIAAITESDK--FFINGSNWEGILGLAYAE 188
DB 131 VVIATGTSMSGILGYDVAV---SSIDVQNIIFGLSETEPGSFYFNCDFGILGLAPPS 187
QY 189 IARPDLSLEPFDFSLVKQTHV-PNLFSLQICGAGFPLNQSEVLASVSGSMIIGDHSLY 247
DB 188 IS--SSGATPEVDNMQSHLVAQDLFVYLSKDG-----ETGSFVLFGGIDPNYT 235
QY 248 TGLSWYTPIRREYVEVLIIVRVEINGDQL--MDCKEYNDKSIDVSGTTLRLPKKVF 305
DB 236 TKGIVWPLSAETVQIOTMDRVTVGNKYVACFTTC-----QAIVDTGSLLLVMPQAYN 289
QY 306 AAVKSIKAASSTEFKPDGFWLGEOLV-CWQAGTTPWNIFFVISLYLMGEVNTQS 358
DB 290 RIKDLGVSSDGEISCDIDSKLPD-----VTFHINGHA-----322
QY 359 FRITILPQYLRPVEDVATSDQCYKFAISQSTGT-----VMGAVIMEGFYVVPDRAR 412
DB 323 -----FTLPASAYVLNEDGSCMLGFENMGTPTELGEOIILGDVFIREFYVIFDRAN 373
QY 413 KRIGFA 418
DB 374 NKVGLS 379

RESULT 9
A34401
C:Species: Homo sapiens (man)
C:Date: 22-Jun-1990 #sequence_revision 22-Jun-1990 #text_change 22-Jun-1999
C:Accession: A42038; A34401; S35663; S34467; A34643
R:Azuma, T.; Liu, W.; Vander Laan, D.J.; Bowcock, A.M.; Taggart, R.T.
J. Biol. Chem. 267, 1609-1614, 1992

A:Title: Human gastric cathepsin E gene. Multiple transcripts result from alternative pd
A:Reference number: A42038; MUID:92112877
A:Accession: A42038
A:Molecule type: DNA
A:Residues: 1-396 <AZU>
A:Cross-references: GB:M82847; NID:g181203; PIDN:AAA52300.1; PID:g181205
A:Note: sequence extracted from NCB1 backbone (NCBIN:75963, NCBIN:75966, NCBIN:75971, NC
R:Azuma, T.; Pals, G.; Mohandas, T.K.; Couvreur, J.M.; Taggart, R.T.
J. Biol. Chem. 264, 16748-16753, 1989
A:Title: Human gastric cathepsin E. Predicted sequence localization to chromosome 1, and
A:Reference number: A34401; MUID:89380302
A:Accession: A34401
A:Molecule type: mRNA
A:Residues: 1-396 <AZ2>
A:Cross-references: GB:J05036; NID:g181193; PIDN:AAA52130.1; PID:g181194
R:Takeda-Ezaki, M.; Yamamoto, K.
Arch. Biochem. Biophys. 304, 352-358, 1993
A:Title: Isolation and biochemical characterization of procathepsin E from human erythro
A:Reference number: S35663; MUID:93349047
A:Accession: S35663
A>Status: preliminary
A:Molecule type: protein
A:Residues: 20-38;54-76 <TAK>
R:Hill, J.; Montgomery, D.S.; Kay, J.
FEBS Lett. 326, 101-104, 1993
A:Title: Human cathepsin E produced in E. coli.
A:Reference number: S34467; MUID:93314762
A:Accession: S34467
A>Status: preliminary
A:Molecule type: protein
A:Residues: 57-60;62-81 <HIL>
R:Ataouda, S.B.P.; Matsuzaki, O.; Kageyama, T.; Takahashi, K.
Biochem. Biophys. Res. Commun. 168, 878-885, 1990
A:Title: Structural evidence for two isozymic forms and the carbohydrate attachment site
A:Reference number: A34643; MUID:90241267
A:Accession: A34643
A>Status: preliminary
A:Molecule type: protein
A:Residues: 54-58,'XXX',62-64,'M',66-89,'X',91-95 <ATH>
A:Accession: B34643
A>Status: preliminary
A:Molecule type: protein
A:Residues: 54-59,'X',61-68 <AT2>
C:Genetics:
A:Gene: GDB:CTSE
A:Cross-references: GDB:119821; OMIM:116890
A:Map position: lq31-lq31
C:Superfamily: pepsin
C:Keywords: aspartic proteinase; blocked amino end; hydrolase; zymogen
F:1-17/Domain: signal sequence #status predicted <SIG>
F:18-53/Domain: activation peptide #status predicted <PRO>
F:54-396/Product: cathepsin E #status predicted <MAT>
F:18/Modified site: blocked amino end (Gln) (in mature form) (probably pyrrolidone carb
F:96,281/Active site: Asp #status predicted

Query Match 11.2%; Score 299.5; DB 2; Length 396;
Best Local Similarity 25.9%; Pred. No. 2.7e-16;
Matches 100; Conservative 65; Mismatches 148; Indels 73; Gaps 16;
QY 50 DEPEEPGRGFSFVENVLRCKSGQGYVVTGKSGEGLTDLVSPHGPNTVTVRANIAAI 165
Db 63 DQSAKEP-----LINVLD-----MEYFGTISGSPQNFVTFDGTSSNLWVPSVYCT 110
QY 106 APHPELHRYQRLSSTYRDLRKGVVPTQKSGEGLTDLVSPHGPNTVTVRANIAAI 165
Db 111 SPACKTHSRFQPSQSTYSQPCQSFIQYGTGSLSGIIGADQVSV-EGLTVVGQFGESV 169
QY 166 TESDKFFINGSNWEGILGLAYAEIARPDSPFFEDSLVKQTHVPLNLFSLQLCGAGFPUN 225
Db 170 TEPGQTQVD-AEFDGILGLGYPSLA--VGGVTPVFDNMMAQ-----NLVDLPWFVYVMSN 222
QY 226 QSEVLASVGGSMIIGDHSLSYTGSLWTPPIRREWYEVIIVRVEINQDLKMDCKEYNY 285

Db 223 PE---SGAGSELIFGGYDHSFSGSLNVPVTKQAYWQIALDNIQVGG--TYMFCSE--G 275
QY 286 DKSIVDSGTTNLRPKKVFEEAAVKSIKAASSTKPEPDGFWLGEQLVCMQAGTTPNIFPV 345
Db 276 COAIVDTGSLTSGDKIKOLONAIGRAP-----VDEYAVE-----CANLNVMPD 322
QY 346 ISLYLMGEVTVNOSFRITLPOQYLRVEDVATSDQDCYKFAISQSSTG----- 393
Db 323 VTFTING-----VPYTLSPAT--TLDFVDMQFC-----SSGFGQLDIHPAGP 366
QY 394 TVMGAVIMEGFYVVDRAKRGIFA 418
Db 367 LWILGDVFIROFYVFDRCNNRVGLA 392
RESULT 10
A41443
pepsin (EC 3.4.23.-) precursor, embryonic - chicken
C:Species: Gallus gallus (chicken)
C:Date: 05-Jun-1992 #sequence_revision 05-Jun-1992 #text_change 21-Jul-2000
C:Accession: A41443
R:Havashi, K.; Agata, K.; Mochii, M.; Yasugi, S.; Eguchi, G.; Mizuno, T.
J. Biochem. 103, 290-296, 1988
A:Title: Molecular cloning and the nucleotide sequence of cDNA for embryonic chicken
A:Reference number: A41443; MUID:88227903
A:Accession: A41443
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-363 <HAV>
A:Cross-references: GB:D00215; NID:g2760810; PIDN:BAA00153.1; PID:g2222853
C:Superfamily: pepsin
C:Keywords: aspartic proteinase; hydrolase; protein digestion

Query Match 11.2%; Score 298.5; DB 2; Length 383;
Best Local Similarity 25.2%; Pred. No. 3.1e-16;
Matches 90; Conservative 75; Mismatches 125; Indels 67; Gaps 14;
QY 77 YVEMTVGSPPTNLIVDTGSSNFAVGA----APHPELHRYQRLSSTYRDLRKGVV 132
Db 76 YGTISIGTPPDFTVVDGTGSSNLWVPSVCTSPACQSHQMFNPSSQSTYKSTGQNLIS 135
QY 133 PYTQKWEGETLTVSPHGPNTVTVRANIAAITESDKFFINGSNWEGILGLAYAEIARP 192
Db 136 HYGTDMEGTGCDTIVTASLMDTNQLFGLST--SEPGQFFVY-VKFDGILGLYPSLAA- 192
QY 193 DLSLEPFDLSVKQTHV-PNLFSLQLCGAGFPNLQSEVLASVGGSMIIGDHSLSYTGSL 251
Db 193 -DGITPVDFNMVNESLLEQNLFVYLS-----REPMGSMVVFQGDISTFTGSI 240
QY 252 WYTPIRREWYEVIIVRVEINQDL--KMDCKEYNYDKSIVDSGTTNLRPKKVFEEAAVK 309
Db 241 NMIPVSYQGYQWISDMSIIVNKEIACSSGC-----QALIDTGTSLVAGPASDINDIQS 294
QY 310 SIKAASTKFPDGFNLGELVCMQAGTTPNIFPVISL-----YLMGEVTVNOSFRITL 365
Db 295 AVGANQNT-----YGEYSV-----NCSHILAMPDVVFGIVGFI----- 326
QY 366 QOYLRPVEDVA---TSQDDCYKFAISQSSTGTVMGAVIMEGFYVVDRAKRGIFA 418
Db 327 -QY--P/PALAYTEQNGQGTCTMSSFNQNSADLWILGDVFIROYYSIFDRANNRVGLA 380
RESULT 11
KHUUD
cathepsin D (EC 3.4.23.5) precursor [validated] - human
N:Alternate names: preprocathepsin D
C:Species: Homo sapiens (man)
C:Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 15-Sep-2000
C:Accession: A25771; S30749; PC2066; I59236; I57716
R:Faust, P.L.; Kornfeld, S.; Chirgwin, J.M.
Proc. Natl. Acad. Sci. U.S.A. 82, 4910-4914, 1985

A:Title: Cloning and sequence analysis of cDNA for human cathepsin D.
A:Reference number: A25771; MUID:85270436
A:Accession: A25771
A:Molecule type: mRNA
A:Residues: 1-412 <FAU>
A:Cross-references: EMBL:M11233; NID:g181179; PIDN:AAB59529.1; PID:g181180
R:Westley, B.R.; May, F.E.B.
Nucleic Acids Res. 15, 3773-3786, 1987
A:Title: Oestrogen regulates cathepsin D mRNA levels in oestrogen responsive human breast
A:Reference number: S30749; MUID:87231068
A:Accession: S30749
A:Molecule type: mRNA
A:Residues: 1-412 <WES>
A:Cross-references: EMBL:X05344; NID:g29677; PIDN:CAA28955.1; PID:g29678
R:May, F.E.B.; Smith, D.J.; Westley, B.R.
Gene 134, 277-282, 1993
A:Title: The human cathepsin D-encoding gene is transcribed from an estrogen-regulated promoter
A:Reference number: PC2066; MUID:94085791
A:Accession: PC2066
A:Molecule type: DNA
A:Residues: 1-23 <NAV>
A:Cross-references: GB:L12980; NID:g291930; PIDN:AAAL6314.1; PID:g455429
A:Experimental source: MCF-7 cell
R:Cavallies, V.; Augereau, P.; Rochefort, H.
Proc. Natl. Acad. Sci. U.S.A. 90, 203-207, 1993
A:Title: Cathepsin D gene is controlled by a mixed promoter, and estrogens stimulate only
A:Reference number: 159236; MUID:93126342
A:Accession: 159236
A>Status: translation not shown; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-22 <CAVI>
A:Cross-references: GB:S52557; NID:g263124; PIDN:AAD1368.1; PID:g4261568
R:Augereau, P.; Miralles, F.; Cavallies, V.; Gaudelot, C.; Parker, M.; Rochefort, H.
Mol. Endocrinol. 8, 693-703, 1994
A:Title: Characterization of the proximal estrogen-responsive element of human cathepsin D
A:Reference number: 157716; MUID:95021301
A:Accession: 157716
A>Status: translation not shown; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-22 <CAVI>
A:Cross-references: GB:S74689; NID:g786350; PIDN:AAD14156.1; PID:g4261856
R:Baldwin, E.T.; Bhat, T.N.; Gulnik, S.; Erickson, J.W.
submitted to the Brookhaven Protein Data Bank, April 1993
A:Reference number: A51839; PDB:1LYA
A:Contents: annotation; X-ray crystallography, 2.5 angstroms, residues 65-161;170-241
R:Baldwin, E.T.; Bhat, T.N.; Gulnik, S.; Erickson, J.W.
submitted to the Brookhaven Protein Data Bank, April 1993
A:Reference number: A51840; PDB:1LYB
A:Contents: annotation; X-ray crystallography, 2.5 angstroms, with inhibitor residues 65-161;170-241
R:Baldwin, E.T.; Bhat, T.N.; Gulnik, S.; Hosur, M.V.; Sowder II, R.C.; Cachau, R.E.; Col
Proc. Natl. Acad. Sci. U.S.A. 90, 6796-6800, 1993
A:Title: Crystal structures of native and inhibited forms of human cathepsin D: implicat
A:Reference number: A48229; MUID:93342076
A:Contents: annotation; X-ray crystallography, 2.5 angstroms
C:Comment: Cathepsin D is a ubiquitous lysosomal proteinase.
C:Comment: In addition to the propeptide, residues 163-168 and 411-412 are proteolytical
C:Comment: The carbohydrate bound to 134-Asn contains a mannose-6-phosphate that is bound
C:Genetics:
A:Gene: GDB:CTSD
A:Cross-references: GDB:120512; OMIM:116840
A:Map position: 11p15.5-11p15.5
C:Function:
A:Description: limited specificity endopeptidase
A:Pathway: intracellular protein degradation
C:Superfamily: peptin
C:Keywords: aspartic proteinase; glycoprotein; hydrolase; lysosome; protein degradation
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-64/Domain: propeptide #status predicted <PRO>
F:65-162,169-410/Product: cathepsin D #status experimental <MAT>
F:267,329-356/Region: phosphotransferase recognition
F:91-160,110-117,286-290,329-366/Disulfide bonds: #status experimental
F:97,295/Active site: Asp #status experimental
F:134,263/Binding site: carboxylate (Asn) (covalent) #status experimental

Query Match 11.1%; Score 296.5; DB 1; Length 412;
Best Local Similarity 26.8%; Pred. No. 5.1e-16;
Matches 117; Conservative 66; Mismatches 163; Indels 91; Gaps 19;
QY 28 IRLPLR-----SGLGAPLGL-----RLPRETDEEPEEPGRGSGFVEMVD 67
Db 22 VRPLKHTSIRMTSEVGSVEDLIAGKPVSKYSAQVPAVE-----GPIPEVLK 72
QY 68 NLRKSGQGYVEMTVGSPPTNINLVDTGSSNFAVGAAPHPEL-----HRYQRLSS 121
Db 73 NYMDAQ---YGEIGTGPQCFVFDVTGSSNLWVPSIHCKLDDIACWLIHHKNSDKSS 129
QY 122 TYRDLRGVVPVTOGKWEGLGTDLSIP-----HGNVTVRANIAAATESDKFF 172
Db 130 TYVKNGTDFDIHVGSGSLSYLSQDTSVPCQSSASSALGGVKEVQVGEATKQPGIT 189
QY 173 INGSNWEIILGLAYAEIARPDOSLEPFDSLVKOTHV-PNLFSLQLCGAGFPLNQSEVLA 231
Db 190 FIAAKFDGILGMAYPRIS--VNNVLPVFDNLMOQKLVQDNIESFYL-----SRDPDA 239
QY 232 SVGSMIIGIDHSLYTGSLWYTPIRREYWEYVLIIVREI-NGODLKMDCKEYNDKSI 290
Db 240 QPGGELMLGTDTSKYKGSLSYLVNTRKAYQWHLDOQVEVASGLTL---CKE--GCEAI 294
QY 291 DSGTTLRLPKKVEAAVKSTKAASSTKFPDGFNLGEQLV-CWQAGTTPWNIPFVISLY 349
Db 295 DTGSLMWGPVDEYRELQKAGVPLIQ-----GEYMIPEKSVT-----LPAITLK 341
QY 350 LMGEVTNQSFRITILPOOYLRPVEDVATSDQDCYKFAISO-----SSTGVMGAVIMEGF 404
Db 342 LGG---KGYKLS--PEDTYLVKSAQKTL--CLSGFMGMDIPPPGGLMLTGLDVFITGRY 393
QY 405 YVFDRAKRGKIGFAVSA 421
Db 394 YTVFDRDNNRVGFAEA 410
RESULT 12
C38302
pepsin (EC 3.4.23.-) II-2/3 precursor - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 14-Jun-1991 #sequence_revision 14-Jun-1991 #text_change 23-Feb-1997
C:Accession: C38302
R:Kageyama, T.; Tanabe, K.; Koiwai, O.
J. Biol. Chem. 265, 17031-17038, 1990
A:Title: Structure and development of rabbit pepsinogens. Stage-specific zymogens, nu
A:Reference number: A38302; MUID:91009127
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-387 <KAG>
A:Cross-references: GB:J05638
C:Superfamily: pepsin
C:Keywords: aspartic proteinase; hydrolase; phosphoprotein; protein digestion

Query Match 11.1%; Score 296; DB 2; Length 387;
Best Local Similarity 26.9%; Pred. No. 5.1e-16;
Matches 97; Conservative 63; Mismatches 135; Indels 66; Gaps 13;
QY 77 YVEMTVGSPPTNINLVDTGSSNFAVGAAPHPE-----LHRYQRLSSYRDLRKG 129
Db 75 YFGTTSIGTTPDFTVFDTSNLSW---PSTYSSSLACALHKRFNEDSTYGTSET 131
QY 130 VVVPVTOGKWEGLGTDLSIPHPGNVTVRANIAAATESDKFFINGSNWEIILGLAYAEI 189
Db 132 LSITVCTGSMTGILGVDIVKVGSIEDTNIQFGLSKTEPSLTLF--APFDGILGLAYPSI 189
QY 190 ARPDOSLEPFDSLVKOTHV-PNLFSLQLCGAGFPLNQSEVLA SVGSMIIGIDHSLYT 248
Db 190 SSSDAT--PVFDNMMNEGLVSDLFVSILSSD-----EKGSLVMFGGDISSYYT 237

RESULT 14
D38302

Biochem. Biophys. Res. Commun. 179, 190-196, 1991

Search completed: September 6, 2001, 16:45:46
Job time: 490 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:51:03 : Search time 72.75 Seconds
(without alignments)
236.845 Million cell updates/sec

Title: US-09-603-713-3
Perfect score: 2663
Sequence: 1 MASMTGGQMGGRSMAGVLP.....CLRLRQHQHDFADDISLLK 503

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2587	97.1	501	1	BACE_HUMAN
2	2319	94.6	501	1	BACE_RAT
3	2484	93.3	501	1	BACE_MOUSE
4	1175	44.1	518	1	BAE2_HUMAN
5	327	12.3	324	1	PEP1_GADMO
6	306	11.5	387	1	PEP1_RABIT
7	304.5	11.4	388	1	PEP4_MACFU
8	302	11.3	367	1	PEP4_CHICK
9	299.5	11.2	396	1	CATE_HUMAN
10	298.5	11.2	383	1	PEPE_CHICK
11	296.5	11.1	412	1	CATD_HUMAN
12	296	11.1	387	1	PEP2_RABIT
13	296	11.1	391	1	CATE_CAVPO
14	295	11.1	387	1	PEP4_RABIT
15	295	11.1	407	1	CATD_RAT
16	291.5	10.9	388	1	PEP2_MACFU
17	286	10.7	387	1	PEP3_RABIT
18	285.5	10.7	388	1	PEPA_HUMAN
19	285.5	10.7	388	1	PEPA_MACMU
20	285.5	10.7	398	1	CATE_RAT
21	285	10.7	398	1	CATD_CHICK
22	284.5	10.7	410	1	CATD_MOUSE
23	283.5	10.6	388	1	PEP1_MACFU
24	281.5	10.6	381	1	CHYM_SHEEP
25	278.5	10.5	386	1	PEPA_PIG
26	277.5	10.4	396	1	CATE_RABIT
27	277.5	10.4	397	1	CATE_MOUSE
28	276.5	10.4	419	1	CARV_CANAL
29	273.5	10.3	381	1	CHYM_BOVIN
30	273	10.3	388	1	PEPE_BOVIN
31	270.5	10.2	377	1	PEPC_MACFU
32	270	10.1	376	1	PAG2_BOVIN
33	268	10.1	396	1	CARP_NEUCR

RESULT 1

ID	BACE_HUMAN	STANDARD	PRT	501 AA.
AC	P56817; Q9UJ75;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DT	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	BETA-SECRETASE PRECURSOR (EC 3.4.23.-) (BETA-SITE APP CLEAVING ENZYME)			
DE	(BETA-SITE AMYLOID PRECURSOR PROTEIN CLEAVING ENZYME) (ASPARTYL PROTEASE 2) (ASP 2) (MEMBRANE-ASSOCIATED ASPARTIC PROTEASE 2)			
DE	(MEMAPIN-2)			
GN	BACE OR BACE1.			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Brain;			
RX	MEDLINE=20002972; PubMed=10531052;			
RA	Vassar R., Bennett B.D., Babu-Khan S., Kahn S., Mendiaz E.A., Denis P., Teplow D.B., Ross S., Anarante P., Loeloff R., Luo Y., Fisher S., Fuller J., Edenson S., Lille J., Jarosinski M.A., Biere A.L., Curran E., Burgess T., Louis J.-C., Collins F., Treanor J., Rogers G., Citron M.;			
RA	"Beta-secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE.";			
RL	Science 286:735-741(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A., SEQUENCE OF 46-68, AND CHARACTERIZATION.			
RC	TISSUE=Brain;			
RX	MEDLINE=20057171; PubMed=10591214;			
RA	Sinha S., Anderson J.P., Barbour R., Basi G.S., Caccavello R., Davis D., Doan M., Dovey H.F., Frigon N., Hong J., Jacobson-Croak K., Jewett N., Keim P., Knops J., Lieberburg I., Power M., Tan H., Tatsuno G., Tung J., Schenk D., Seubert P., Suomensaar S.M., Wang S., Walker D., Zhao J., McConlogue L., Varghese J.;			
RA	"Purification and cloning of amyloid precursor protein beta-secretase from human brain.";			
RL	Nature 402:537-540(1999).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20057170; PubMed=10591213;			
RA	Yan R., Blenkowski M.J., Shuck M.E., Miao H., Tory M.C., Pauley A.M., Brashers J.R., Stratan N.C., Mathews W.R., Buhl A.E., Carter D.B., Tomasselli A.G., Patodi L.A., Heinrich R.L., Gurney M.E.;			
RA	"Membrane-anchored aspartyl protease with Alzheimer's disease beta-secretase activity.";			
RL	Nature 402:533-537(1999).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20030166; PubMed=10561122;			
RA	Hussain I., Powell D.J., Howlett D.R., Tew D.G., Meek T.D., Chapman C., Gloger I.S., Murphy K.E., Southan C.D., Ryan D.M., Smith T.S., Simmons D.L., Walsh F.S., Dingwall C., Christie G.;			
RA	"Identification of a novel aspartic proteinase (Asp 2) as beta-			

34	266	10.0	405	1	CARP_YEAST
35	263	9.9	394	1	PEPC_CAVPO
36	261.5	9.8	387	1	ASPP_AEDAE
37	261.5	9.8	388	1	PEPC_HUMAN
38	261	9.8	388	1	PAG_HORSE
39	260	9.8	345	1	CATD_PIG
40	258.5	9.7	496	1	ASPR_ORYSA
41	253	9.5	392	1	PEPC_RAT
42	251.5	9.4	389	1	PAG1_PIG
43	251.5	9.4	509	1	APR1_ORYSA
44	250	9.4	402	1	RENI_RAT
45	250	9.4	406	1	RENI_HUMAN

ALIGNMENTS

secretase.";
 RL Mol. Cell. Neurosci. 14:419-427(1999).
 RN
 RP SEQUENCE OF 14-501 FROM N.A., AND CHARACTERIZATION.
 RX MEDLINE=20144060; PubMed=10677483;
 RA Lin X., Koelsch G., Wu S., Downs D., Dashti A., Tang J.;
 RT "Human aspartic protease memapsin 2 cleaves the beta-secretase site of
 beta-amyloid precursor protein".
 RL Proc. Natl. Acad. Sci. U.S.A. 97:1456-1460(2000).
 CC
 CC -!- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC PROCESSING OF THE
 CC AMYLOID PRECURSOR PROTEIN (APP). CLEAVES AT THE AMINO TERMINUS OF
 CC THE A-BETA PEPTIDE SEQUENCE, BETWEEN RESIDUES 671 AND 672 OF APP,
 CC LEADS TO THE GENERATION AND EXTRACELLULAR RELEASE OF BETA-CLEAVED
 CC SOLUBLE APP, AND A CORRESPONDING CELL-ASSOCIATED CARBOXY-TERMINAL
 CC FRAGMENT WHICH IS LATER RELEASE BY GAMMA-SECRETASE.
 CC
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC
 CC -!- TISSUE SPECIFICITY: BRAIN.
 CC
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
 CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF190725; AAF04142.1; -
 CC EMBL; AF201468; AAF18982.1; -
 CC EMBL; AF200343; AAF17079.1; -
 CC EMBL; AF204943; AAF26367.1; -
 CC EMBL; AF200193; AAF13715.1; -
 CC MTM; 604252; -
 CC InterPro: IPR001461; -
 CC Pfam; PF00026; asp; 3
 CC PRINTS; PR00792; PEPsin.
 CC PROSITE; PS00141; ASP_PROTEASE; 1.
 CC Hydrolase; Aspartyl protease; Glycoprotein; zymogen; Transmembrane;
 CC Signal.
 CC SIGNAL 1 21 POTENTIAL.
 CC PROPEP 22 45
 CC CHAIN 46 501
 CC DOMAIN 22 457
 CC TRANSMEM 458 478
 CC DOMAIN 479 501
 CC ACT SITE 93 93
 CC ACT SITE 289 289
 CC CARBOHYD 153 153
 CC CARBOHYD 172 172
 CC CARBOHYD 223 223
 CC CARBOHYD 354 354
 CC SEQUENCE 501 AA; 55763 MW; 377CE4C824ACEF05 CRC64;
 SQ
 Query Match 97.1%; Score 2587; DB 1; Length 501;
 Best Local Similarity 100.0%; Pred. No. 9.8e-198;
 Matches 488; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 16 AGVLPAGHTQHGIRLPLRSGLGAPLGLRLPRETDEPEEPGRGSGVEMVDNLGRKSGQ 75
 DB 14 AGVLPAGHTQHGIRLPLRSGLGAPLGLRLPRETDEPEEPGRGSGVEMVDNLGRKSGQ 73
 QY 76 GYIVEMTVGSPQTNLNLVDGTSSNFVGAAPHLRHYQRLSSVYRDLRKGVVYPYT 135
 DB 74 GYIVEMTVGSPQTNLNLVDGTSSNFVGAAPHLRHYQRLSSVYRDLRKGVVYPYT 133
 QY 136 QKWEGETLTLVSPHGNVTVRANIAAITESDKFFTINGSNWEGILGLAYAEIARPDPS 195
 DB 134 QKWEGETLTLVSPHGNVTVRANIAAITESDKFFTINGSNWEGILGLAYAEIARPDPS 193
 QY 196 LEFFFDLSLVKQTHVPLNLSQLCGAGFLNQSEVLASVGGSMIIGIDHSLYTGSLWYTP 255
 DB 196 LEFFFDLSLVKQTHVPLNLSQLCGAGFLNQSEVLASVGGSMIIGIDHSLYTGSLWYTP 255

DB 194 LEFFFDLSLVKQTHVPLNLSQLCGAGFLNQSEVLASVGGSMIIGIDHSLYTGSLWYTP 253
 QY 256 IRREWYEVIIIVRVEINGQDLKMDCKEYNDKSIYDSTTNLRPKKVFEEAAVKSIAAS 315
 DB 254 IRREWYEVIIIVRVEINGQDLKMDCKEYNDKSIYDSTTNLRPKKVFEEAAVKSIAAS 313
 QY 316 STEKEPDCFWLGEQLVVCWAGTTPWNIFFVLSLYLMGEVTVNQSFRTITLPOQYLRPVEDV 375
 DB 314 STEKEPDCFWLGEQLVVCWAGTTPWNIFFVLSLYLMGEVTVNQSFRTITLPOQYLRPVEDV 373
 QY 376 ATSQDDCYKFAISQSSTGTVMGAVIMEGFYVYVDFRARRIGRIGFVNSACHVDEFTAAVEG 435
 DB 374 ATSQDDCYKFAISQSSTGTVMGAVIMEGFYVYVDFRARRIGRIGFVNSACHVDEFTAAVEG 433
 QY 436 PFVTLDMDCGYNIPQTDSTLTMTIAYVMAAICALEMPLCLMVCWQRCRLCRLQHQHDF 495
 DB 434 PFVTLDMDCGYNIPQTDSTLTMTIAYVMAAICALEMPLCLMVCWQRCRLCRLQHQHDF 493
 QY 496 ADDISLLK 503
 DB 494 ADDISLLK 501
 RESULT 2
 BACE-RAT
 ID P56819; STANDARD; PRT; 501 AA.
 AC 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE BETA-SECRETASE PRECURSOR (EC 3.4.23.-) (BETA-SITE APP CLEAVING ENZYME)
 DE (BETA-SITE AMYLOID PRECURSOR PROTEIN CLEAVING ENZYME) (ASPARTYL
 DE PROTEASE 2) (ASP 2) (ASP2) (MEMBRANE-ASSOCIATED ASPARTIC PROTEASE 2)
 DE (MEMAPSIN-2).
 GN BACE
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20002972; PubMed=10531052;
 RA Vassar R., Bennett B.D., Babu-Khan S., Kahn S., Mendiaz E.A.,
 RA Denis P., Teplow D.B., Ross S., Amarante P., Loeloff R., Luo Y.,
 RA Fisher S., Fuller J., Edenson S., Lile J., Jarosinski M.A.,
 RA Biere A.L., Curran E., Burgess T., Louis J.-C., Collins F.,
 RA Treanor J., Rogers G., Citron M.;
 RT "Beta-secretase cleavage of Alzheimer's amyloid precursor protein by
 the transmembrane aspartic protease BACE".
 RL Science 286:735-741(1999).
 CC -!- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC PROCESSING OF THE
 CC AMYLOID PRECURSOR PROTEIN (APP). CLEAVES AT THE AMINO TERMINUS OF
 CC THE A-BETA PEPTIDE SEQUENCE, BETWEEN RESIDUES 671 AND 672 OF APP,
 CC LEADS TO THE GENERATION AND EXTRACELLULAR RELEASE OF BETA-CLEAVED
 CC SOLUBLE APP, AND A CORRESPONDING CELL-ASSOCIATED CARBOXY-TERMINAL
 CC FRAGMENT WHICH IS LATER RELEASE BY GAMMA-SECRETASE (BY
 CC SIMILARITY).
 CC
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
 CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF190727; AAF04144.1; -
 CC InterPro: IPR001461; -
 CC Pfam; PF00026; asp; 3
 CC PRINTS; PR00792; PEPsin.

Fri Sep 7 10:58:21 2001

```

Db      14  SCMLPAQOHTLGLRPLRSLGAGPLGLRLPRETDESEFGRGSRFVMDNLKSGSQ 73
Qy      76  GYVEMTVGSPQTINILVDGTSSNFAVGAAPHPFLHRYVQROSLSSYRDLRGVYVPT 135
Db      74  GYVEMTRQOPLTKNLILVDGTSSNFAVGAAPHPFLHRYVQROSLSSYRDLRGVYVPT 133
Qy      136  QKWEGELGTLVSPHGPNTVTRANIAITSDKFFINGSNWEGILGLAYAEIARPDSD 195
Db      134  QKWEGELGTLVSPHGPNTVTRANIAITSDKFFINGSNWEGILGLAYAEIARPDSD 193
Qy      196  LEPPFDLVKQTHVNLFSQLCGAGFLNQLSEVLASVGGSMIIGGDHSLYTGSLWYTP 255
Db      194  LEPPFDLVKQTHVNLFSQLCGAGFLNQLSEVLASVGGSMIIGGDHSLYTGSLWYTP 253
Qy      256  IRREWEYEVIIIVRVEINGDLKMDCKEYNDKSIYDSTGNLRLPKVFEAAVKSIRAS 315
Db      254  IRREWEYEVIIIVRVEINGDLKMDCKEYNDKSIYDSTGNLRLPKVFEAAVKSIRAS 313
Qy      316  STEKFDGFWLGEOLYCWQAGTTPNIFPVISLYLMGEVNTQSFRTITLPOOYLRPVEDV 375
Db      314  STEKFDGFWLGEOLYCWQAGTTPNIFPVISLYLMGEVNTQSFRTITLPOOYLRPVEDV 373
Qy      376  ATSDCCYKFAISQSSTGTVMGAVIMEGYVDFDRKRGIGFAVSNCHVHDEFTAAVEG 435
Db      374  ATSDCCYKFAISQSSTGTVMGAVIMEGYVDFDRKRGIGFAVSNCHVHDEFTAAVEG 433
Qy      436  PFVTLMDKCGYNIPODESLMTIAYVMAAICALFMLPCLMVCOWRCRLCRLRQHDDF 495
Db      434  PFVTLMDKCGYNIPODESLMTIAYVMAAICALFMLPCLMVCOWRCRLCRLRQHDDF 493
Qy      496  ADDISLLK 503
Db      494  ADDISLLK 501

RESULT 4
BAE2_HUMAN
ID BAE2_HUMAN STANDARD; PRT; 518 AA.
AC Q9Y520; Q9UJT6;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE BETA SECRETASE 2 PRECURSOR (EC 3.4.23.-) (BETA-SITE APP-CLEAVING
DE ENZYME 2) (ASPARTYL PROTEASE 1) (ASP 1) (ASP1) (MEMBRANE-ASSOCIATED
DE DE ASPARTIC PROTEASE 1) (MEMAPSN-1).
GN BACE2 OR ASP21.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=20057170; PubMed=10591213;
RA Yan R., Bienkowski M.J., Shuck M.E., Miao H., Tory M.C., Pauley A.M.,
RA Braslier J.R., Stratman N.C., Mathews W.R., Buhl A.E., Carter D.B.,
RA Tomaselli A.G., Parodi L.A., Heinrichson R.L., Gurney M.E.;
RA "Membrane-anchored aspartyl protease with Alzheimer's disease
RT beta-secretase activity.";
RL Nature 402:533-537(1999).
RN [2]
SEQUENCE FROM N.A.
RC TISSUE=Bone marrow.
RA Xin H., Stephens J.C., Duan X., Harrowe G., Kim E., Grieshammer U.,
RA Glese K.;
RT "Identification of a novel aspartic-like protease differentially
RT expressed in human breast cancer cell lines.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
SEQUENCE FROM N.A.
RP Accarino M.P., Funagalli P., Ottolenghi S., Taramelli R.;
RA "Cloning of a gene from chromosome 21 Down region encoding a potential
RT transmembrane aspartyl protease.";

```

Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.

SEQUENCE FROM N.A.

Solans A., Estivill X., de la Luna S.;

"Cloning of a novel mammalian aspartyl protease.";

Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.

SEQUENCE FROM N.A.

MEDLINE=20030166; PubMed=10561122;

Hussain I., Powell D.J., Howlett D.R., Tew D.G., Meek T.D.,

Chapman C., Gloger I.S., Murphy K.E., Southan C.D., Ryan D.M.,

Smith T.S., Simmons D.L., Walsh F.S., Dingwall C., Christie G.;

"Identification of a novel aspartic proteinase (Asp 2) as

beta-secretase.";

Mol. Cell. Neurosci. 14:419-427(1999).

SEQUENCE FROM N.A.

MEDLINE=20144060; PubMed=10677483;

Lin X., Koelsch G., Wu S., Downs D., Dashti A., Tang J.;

"Human aspartic protease memapsin 2 cleaves the beta-secretase site of

beta-amyloid precursor protein.";

Proc. Natl. Acad. Sci. U.S.A. 97:1456-1460(2000).

SEQUENCE FROM N.A.

MEDLINE=20289799; PubMed=10830953;

Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,

Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,

Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,

Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,

Reichwald K., Rump A., Schillhaber M., Schudy A., Zimmermann W.,

Rosenchal A., Sasaki T., Nagamine K., Kawasaki K., Asakawa S.,

Shintani A., Shimizu N., Nordsiek G., Hornischer K., Antonarakis S.E.,

Minoshima S., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,

Scharfe M., Beck A., Klages S., Hennig S., Riessmann L., Dagand E.,

Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,

Lehrach H., Reinhardt R., Yaspo M.-L.;

"The DNA sequence of human chromosome 21.";

Nature 405:311-319(2000).

-!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE

EUKARYOTIC ASPARTYL PROTEASES FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF200342; AAF17078.1;

EMBL; AF117892; AAD45240.1;

EMBL; AF050171; AAD45963.1;

EMBL; AF178532; AAF29494.1;

EMBL; AF204944; AAF26368.1;

EMBL; AF200192; AAF13714.1;

EMBL; AL163284; CAB90458.1;

EMBL; AL163285; CAB90554.1;

InterPro; IPR001461;

Pfam; PF00026; asp; 3.

PRINTS; PR00792; PEPsin.

PROSITE; PS00141; ASP_PROTEASE; 2.

Hydrolase; Aspartyl protease; Glycoprotein; Zymogen; Transmembrane;

Signal.

Signal. 1 20 POTENTIAL.

FT PROPEP 21 ? POTENTIAL.

FT CHAIN ? 518 BETA SECRETASE 2.

FT DOMAIN 21 473 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 474 494 POTENTIAL.

FT DOMAIN 495 518 CYTOPLASMIC (POTENTIAL).

FT ACT_SITE 110 110 BY SIMILARITY.

FT ACT_SITE 303 303 BY SIMILARITY.


```
FT CARBOHYD 170 170 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 366 366 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 36 36 A -> T (IN REF. 6).
SQ SEQUENCE 518 AA; 56180 MW; 2E903150823760D3 CRC64;

Query Match 44.1%; Score 1175; DB 1; Length 518;
Best Local Similarity 47.1%; Pred. No. 1.1e-85;
Matches 228; Conservative 81; Mismatches 151; Indels 24; Gaps 5;

QY 20 PAHGTHQIRGLPLRSGLGAPLGLRLPRETDEPEPGRGRGSEVEMVNDLRCKSGGYV 79
DB 52 PAERHADGLALEPALA-----SPAGAAFLAMVNDLQDGSGRGYL 94
QY 80 EMTVSGPPQTLNLVDTGSSNFAVGAAPHPFLHRYRQRLSSYRDLRKGVVVYPTQGW 139
DB 95 EMLIGTPPKQLIYDTGSSNFAVAGTPHSYIDTYDTERSTYRSKGFDTVKYQTQGSW 154
QY 140 EGEGLGDLVSIPIHGNVTVRANIAATESDKFFINGSNNEGILGLAYAEIARPDSDLEPF 199
DB 155 TGFVGEDLVTPKGFNTSFLVNATIFESNFPLPGIKWNGILGLAYATLAKPSSLET 214
QY 200 FDSLVRQTHVNLFSLOLCAGFPLNQSEVLASVCGSMIIGGDHSLYTGSLWYTPIRRE 259
DB 215 FDSLVTQANIPNVFSMOMGAGLPVAGS-----GTNGGSLVLGGIEPSLYKGDWYTPIKEE 271
QY 260 WYVEIIVRVEINGODLMDCKEYNDKSIDVSGTTLNLRPKKVFEEAAVKSIAKASTEK 319
DB 272 WYQIEILKLEIGGQSLNLDREYNADKAIVDSGTTLLRLPKQVFDVAVAEVARASLIPE 331
QY 320 FPDGFWLGOLVCMQAGTTPWNIFFVISYLMGEVNTNQSFRTILPQOYLPRVEDVATSO 379
DB 332 FSDGFWTGSQACWTNSETPMWSYFPRKISYLRDENSRSFRITILPOLYIQPMGAGLNY 391
QY 380 DDCYKFAISQSSTGTVMGAVIMEGFFVVDRAKRGTSVACHVDFERTAAVEGPFVT 439
DB 392 -ECYREGISFSPNALVIGATVMEGFFVIFDRAKRGVFAASPCAETAAGAVSISGPFST 450
QY 440 LDMEDCGYINPQDTESLMTIAYVMAAIC-ALFMPLCLMVCWMCRGLRQOHDDFAD 498
DB 451 EDVASNCVPAQSILSEPLWISVYALMSVCGAILLVLLVLLLPFCRQ--RPRDPEVND 508
QY 499 ISLL 502
DB 509 ESSL 512

RESULT 5
PEPL_GADMO STANDARD; PRT; 324 AA.
AC P56272;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PEPSIN IIB (EC 3.4.23.-).
OS Gadus morhua (Atlantic cod).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadoidei; Gadidae;
OC Gadus.
OX NCBI_TaxID=8049;
RN [1]
RP SEQUENCE, AND X-RAY CRYSTALLOGRAPHY.
RC TISSUE=Stomach;
RA Karlsten S., Hough E., Olsen R.L.;
RT "Structure and proposed amino-acid sequence of a pepsin from Atlantic
RT cod (Gadus morhua).";
RL Acta Crystallogr. D 54:32-46(1998).
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
DR PDB; 1AM5; 24-DEC-97.
DR InterPro; IPR001461; -.
DR InterPro; IPR001969; -.
```

```
DR Pfam; PF00026; asp; 1.
DR PRINTS; PR00792; PEPsin.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
KW Hydrolase; Aspartyl protease; Digestion; Stomach; Gastric juice;
KW 3D-structure. 32 32 BY SIMILARITY.
FT ACT_SITE 32 32 BY SIMILARITY.
FT ACT_SITE 214 214 BY SIMILARITY.
FT DISULFID 45 50 BY SIMILARITY.
FT DISULFID 206 209 BY SIMILARITY.
FT DISULFID 247 280 BY SIMILARITY.
SQ SEQUENCE 324 AA; 34014 MW; EE3A6097B6941DD7 CRC64;

Query Match 12.3%; Score 327; DB 1; Length 324;
Best Local Similarity 27.9%; Pred. No. 1.1e-18;
Matches 104; Conservative 66; Mismatches 137; Indels 66; Gaps 15;

QY 65 MYDNLRGKSGQYVYEMTVGSPPTNLNLVDTGSSNFAVG-----AAPHFLHRYRQRLS 120
DB 2 VTEQMKNEADTEYGVISIGTTPSEFKVIFDTGSSNLWVSSSHCSAQACSNHNKFKPROS 61
QY 121 STYRDLRKGVYPTQGWEGELGTLVSIPIHG--PNVTVRANIAAITESDKFFINGSNW 178
DB 62 STYVEGKTVDLTYGTGGMRLGQDTSVGGSDPNQELG---ESQTEGPFQA-AAFP 117
QY 179 EGILGLAYAEIARPDSDLEPFDFSLVKQTHV-PNLSLOLCGAGFPLNSEVLASVCGSM 237
DB 118 DGIILGLAYSIAAA--GAVPVEDNMGSSQLVKDLFSFYLSSGG--ANGSEVN----- 166
QY 238 IIGGIDHSLYTGSLWYTPIRREWYEVIIIVRVEINGODLMD-CKEYNDKSIDVSGTTLN 296
DB 167 -LGGVDNSHYTSGIHWIPVTAEKYMQVALDGTITVNGTAAACEGC-----QAIVDTGTSK 219
QY 297 LRLPKVFEEAAVKSIAKASTEKFPDGFMLGEOLVCMQAGTTPWNIFFVISYLMGEVNTN 356
DB 220 IVAPVSALANIKMDIGASEN-----QGEHMGN---CASVQSLPDIIF-----TI 260
QY 357 QSFRTILPQOYLPRVEDVATSOODCYKFAISQSSTGT-----VMGAVIMEGFFVYV 408
DB 261 NGVKQPLPPSAVIEGDOAFCTS-----GLGSSGVPSNTSELWIFGDFLNRNYTIIY 311
QY 409 DRARKIRIGFAVSA 421
DB 312 DRTNNKVGFAVSA 324

RESULT 6
PEPL_RABIT STANDARD; PRT; 387 AA.
AC P28712;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PEPSIN II-2 PRECURSOR (EC 3.4.23.1) (PEPSIN A).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91009127; PubMed=2129536;
RA Kageyama T., Tanabe K., Koizumi O.;
RT "Structure and development of rabbit pepsinogens. Stage-specific
RT zymogens, nucleotide sequences of cDNAs, molecular evolution, and
RT gene expression during development.";
RL J. Biol. Chem. 265:17031-17038(1990).
CC -I- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
CC ALSO CLEAVED TO SOME EXTENT.
CC -I- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.
CC -I- DEVELOPMENTAL STAGE: PEPSINOGENS IN GROUP I, II, AND III WHERE
CC THE PREDOMINANT ZYMOGENS AT LATE POSTNATAL STAGE.
CC -I- MISCELLANEOUS: THE EXPRESSION OF PEPSINOGEN GENES IS REGULATED BY
```

```

CC      HORMONES AND RELATED SUBSTANCES.
CC      -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC      EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC      PIR; B38302; B38302.
CC      DR MEROPS; A01.001; -.
CC      DR InterPro: IPR001461; -.
CC      DR InterPro: IPR001969; -.
CC      DR Pfam: PF00026; asp; 1.
CC      DR PRINTS; PR00792; PEPsin.
CC      DR PROSITE; PS00141; ASP_PROTEASE; 2.
CC      KW Hydrolase; Aspartyl protease; Digestion; Stomach; Gastric juice;
CC      MW Zymogen; Signal; Phosphorylation; Multigene family.
CC      FT SIGNAL 1 15
CC      FT PROPEP 16 59
CC      FT CHAIN 60 387
CC      FT MOD_RES 129 129
CC      FT ACT_SITE 93 93
CC      FT ACT_SITE 276 276
CC      FT DISULFID 106 111
CC      FT DISULFID 267 271
CC      FT DISULFID 310 343
CC      SEQUENCE 387 AA; 42070 MW; A6EC48F715541A48 CRC64;

Query Match 11.5%; Score 306; DB 1; Length 387;
Best Local Similarity 27.1%; Pred. No. 6.4e-17;
Matches 98; Conservative 67; Mismatches 131; Indels 66; Gaps 15;

QY 77 YKVEVTGPPQTLNVLDTGSSNFVAVG-----AAPHFLHRYQRLSTYRDLRKGVYV 132
   | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
DB 75 YFCTISGTPPQEFVIFDTGSSNLWVPSTYCYSLACLKRFNPDDSTFQATSETLSI 134

QY 133 PTQGWKEGELGDLVSPHGNVTVVRANIAITSD---KEFFINGSNWEGILGAYAEI 189
   | | | | | | : : : : | : : : : | : : : : | : : : : | : : : : |
DB 135 TYGTGSGMTGILGYDTVKV---GNIEDTQIFGLSKTEPGITELV--APFDGILGLAYPSI 189

QY 190 ARPDDSLPEFDFSLVKQTHV-PNLFSLQCGAGFLPNQSLVLSVGGSMIIGIDHSLYT 248
   | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
DB 190 SASDAT--PVFNDMNEGLVSEDLFSVLSNG-----EKGSMVFMGIDSSYIT 237

QY 249 GSLWYTPIRREWYEVIIIVRVEINGQDLAM--DKKEYNVKSIDVSGTNNLKLPPKVEA 306
   | | | : : : : : | : : : : | : : : : | : : : : | : : : : |
DB 238 GSLNWVPVSHGQWQITMDSITNGETIACADSC-----QAVVDTGTSLLAGPTSAISK 291

QY 307 AVKSIKAASSTKFPDGFVLGQLV-CWQAGTPWNIPFVISLYLMGEVTVNQSFRITILP 365
   | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
DB 292 IOSYIGASKNL-----LGENIISCSAIDSLPDIVF-----TINN 325

QY 366 QOYLREVED-VATSDQDC---YKFAISQSSTGT--VMGAVIMEGFVYVDFDRARKRIGFAV 419
   | | | : : : : : | : : : : | : : : : | : : : : | : : : : |
DB 326 VOYPLPASAYILKEDDDCLSGDFGMNLDTSYGLWILGDFVIRQVFTVPDRANNQVGLAA 385

QY 420 SA 421
DB 386 AA 387

```

RESULT 7

```

ID PEP4_MACFU STANDARD; PRT; 388 AA.
AC P27678;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE PEPsin A-4 PRECURSOR (EC 3.4.23.1) (PEPSIN I/II).
GN pCa.
OS Macaca fuscata fuscata (Japanese macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9543;
OR [1]

```

```

RP SEQUENCE FROM N.A., AND SEQUENCE OF 16-70.
RC TISSUE=Gastric mucosa;
RX MEDLINE=92037645; PubMed=1935977;
RA Kageyama T., Tanabe K., Koizumi O.;
RT "development-dependent expression of isozymes of monkey
RL Eur. J. Biochem. 202;205-215(1991).
CC -1- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
CC ALSO CLEAVED TO SOME EXTENT.
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-,
CC -1- MISCELLANEOUS: THE EXPRESSION OF PEPSINOGEN GENES IS REGULATED BY
CC HORMONES AND RELATED SUBSTANCES.
CC -1- MISCELLANEOUS: EACH PEPSINOGEN IS CONVERTED TO CORRESPONDING
CC PEPSIN AT PH 2.0 IN PART AS A RESULT OF THE RELEASE OF A 47 AA
CC ACTIVATION SEGMENT AND IN PART AS A RESULT OF STEPWISE PROTEOLYTIC
CC CLEAVAGE VIA AN INTERMEDIATE FORM(S).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X59753; CAA2425.1; -.
DR PIR; S16065; S16065.
DR PIR; S19682; S19682.
DR HSSP; P00790; 1PSO.
DR MEROPS; A01.001; -.
DR InterPro: IPR001461; -.
DR InterPro: IPR001969; -.
DR Pfam; PF00026; asp; 1.
DR PRINTS; PR00792; PEPsin.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
KW Hydrolase; Aspartyl protease; Digestion; Stomach; Gastric juice;
KW Zymogen; Multigene family; Signal; Glycoprotein.
FT SIGNAL 1 15
FT PROPEP 16 38
FT PROPEP 39 62
FT CHAIN 63 388
FT ACT_SITE 94 94
FT ACT_SITE 277 277
FT DISULFID 107 112
FT DISULFID 268 272
FT DISULFID 311 344
FT CARBOHYD 88 88
FT SEQUENCE 388 AA; 41955 MW; A2923AB1F7FCDE9 CRC64;

Query Match 11.4%; Score 304.5; DB 1; Length 388;
Best Local Similarity 27.6%; Pred. No. 8.4e-17;
Matches 108; Conservative 64; Mismatches 136; Indels 83; Gaps 17;

QY 46 PRETDEPEEPGRGSRFVEMVNLKRGSGQGYVEMTVGSPQTLNVLDTGSSNFVAVG 105
   | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
DB 60 PTLDEQPLE-----NYLDV-----EYFGTIGTGAQNFVTVVFDTGSSNLWV-- 102

QY 106 APHPFL-----HRYXORLSSTYRDLRKGVYVPTGKWEGLGTLVSLPHGNVTV 158
   | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
DB 103 -PSVYCYSLACMDHNLFPQDSSSTRATSKTVSYTYGTGSMTGILGYDTVKV---GGISD 158

QY 159 RANIAAITESDK-FFINGSNWEGILGAYAEIARPDSDLPEFDFSLVKQTHV-PNLFSLQ 216
   | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
DB 159 TNQIFGLSETEPGFFLYFAPDGLGLAYPSIS--SSGATRVFDNIWQRLVSDQLFSYV 216

QY 217 LCGAGFLPNQSEVLASVGGSMIIGIDHSLYTGLWYPIREWYEVIIIVRVEINGQDL 276
   | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
DB 217 LSAD-----DQS-----GSWVIFGIDSSYITGSLNWVPVSVGEGWQIVSDSITWNGKI 266

```

```
QY 277 --KMDCKEYNDKSIYDSTGTLNLRPKVFEAAVKSIKAASSTKPKPDGFWLGEQLV-CW 333
Db 267 ACAKGC-----QAIVDTGTSLLTGTPSIANTOSDIGNESD-----GEMVSCS 312
QY 334 QAGTTTNNFFPVLISLMLGEVNTNQSFRIILPQOY-LRPVEDVATSDQDCYK-----PAI 387
Db 313 ASSLSPDIVF-----TINGVQYPLPPSAYILQSGSGTSGTSGFQGMVDP 354
QY 388 SQSSTGTVMGAVIMEGFYVVFDRARRKRGFA 418
Db 355 TESGELWILGDVFIQOYFVFDRAANNQVGLA 385

RESULT 8
PEPA_CHICK
ID PEPA_CHICK STANDARD; PRT; 367 AA.
AC P00793;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PEPIN A PRECURSOR (EC 3.4.23.1).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
ON NCBI_TaxID=9031;
RX MEDLINE=84004412; PubMed=6617663;
RA Baudys M., Kostka V.;
RT "Covalent structure of chicken pepsinogen.";
RL Eur. J. Biochem. 136:89-99(1983).
CC -!- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY. ALTHOUGH BONDS
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
CC ALSO CLEAVED TO SOME EXTENT.
CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
DR PIR; A00984; PECH.
DR HSSP; P00794; 3CMS.
DR MEROPS; A01.001; -.
DR InterPro; IPR001461; -.
DR InterPro; IPR001989; -.
DR Pfam; PF00026; asp. 1.
DR PRINTS; PR00792; PEPIN.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
KW Hydrolase; Aspartyl protease; Digestion; Stomach; Zymogen;
KW Glycoprotein; Gastric juice.
FT PROPEP 1 42 ACTIVATION PEPTIDE.
FT CHAIN 43 367 PEPIN A.
FT ACT_SITE 77 77
FT ACT_SITE 260 260
FT CARBOHYD 113 113 N-LINKED (GLCNAC. . .).
FT DISULFID 90 95
FT DISULFID 251 255
FT DISULFID 290 323
SQ SEQUENCE 367 AA; 40431 MW; 0C547E7FD8F5B341 CRC64;
```

Query Match 11.3%; Score 302; DB 1; Length 367;
Best Local Similarity 24.0%; Pred. No. 1.2e-16;
Matches 88; Conservative 69; Mismatches 125; Indels 84; Gaps 13;

```
QY 77 YVEMTVGSPPTLNLVDVNGSSNFAVGAPHPFL-----HRYVQRQLSSTYRDLRKG 129
Db 59 YGTIGTGPQDFSVIFDGSNLMW---PSIYKSSACSNHRRFDPSPKSSSTVSTNET 115
QY 130 VYVYPTQGWKEGLGDLVSIHPGNVTVVRANIAATESDK-PFINGSNMGILGLAYAE 188
Db 116 VYIAYTGSGNSGILGYDTAV---SSIDVQNIQFGLSETEPGSFYFCNFDGILGLAFPS 172
QY 189 IARPDOSLEPFFSLVKQTHV-PNLSLQLCGAGFPLNQSEVLASVCGSGMIIGIDHSLY 247
| : ||| : | | : ||| : | | : ||| : |
```

```
Db 173 IS--SSGATPVFVNNMNSOHLVAODLFSVYLSKDG-----ETGSFVLFGGIDPNYT 220
QY 248 TGSLSWTPTRREWYEVIIIVRVEINGODLK--MDCKEYNDKSIYDSTGTLNLRPKVFE 305
Db 221 TKGIYVPLSAETIYWOITMDRVTVGNKYACFTC-----QAIVDTGTSLLVMPOGAYN 274
QY 306 AAVKSTKAASSTE-----KFPDGFMLGEQLVCMQAGTTPNFIFFVISLYLMGEVNTNOS 358
Db 275 RIIKDIGVSSDGEISCDIDSKLPD-----VTFFHNGHA----- 307
QY 359 FRITILPQOYLRPEVDVATSDQDCYKFAISQSSTGT-----VMGAVIMEGFYVVFDRAR 412
Db 308 -----FTLPASAYVLNEDGSCMLGFENMGCTPELGEQWILGDVFIQOYFVFDRA 358
QY 413 KRIGFA 418
Db 359 NKVGLS 364

RESULT 9
CATE_HUMAN
ID CATE_HUMAN STANDARD; PRT; 396 AA.
AC P14091;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CATHEPSIN E PRECURSOR (EC 3.4.23.34).
ON CTSE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RX MEDLINE=89380302; PubMed=2674141;
RA Azuma T., Fals G., Mohandas T.K., Couvreur J.M., Taggart R.T.;
RT "Human gastric cathepsin E gene. Multiple transcripts result from
RT alternative polyadenylation of the primary transcripts of a single
RT gene locus at 1q31-q32.";
RL J. Biol. Chem. 267:1609-1614(1992).
RN [3]
RX MEDLINE=92112877; PubMed=1370478;
RA Azuma T., Liu W.G., Vander Laan D.J., Bowcock A.M., Taggart R.T.;
RT "Human gastric cathepsin E gene. Multiple transcripts result from
RT alternative polyadenylation of the primary transcripts of a single
RT gene locus at 1q31-q32.";
RL J. Biol. Chem. 267:1609-1614(1992).
RN [2]
RX MEDLINE=92112877; PubMed=1370478;
RA Azuma T., Liu W.G., Vander Laan D.J., Bowcock A.M., Taggart R.T.;
RT "Human gastric cathepsin E gene. Multiple transcripts result from
RT alternative polyadenylation of the primary transcripts of a single
RT gene locus at 1q31-q32.";
RL J. Biol. Chem. 267:1609-1614(1992).
RX MEDLINE=902411267; PubMed=2334440;
RA Athauda S.B.P., Matsuzaki O., Kgevama T., Takahashi K.;
RT "Structural evidence for two isozymic forms and the carbohydrate
RT attachment site of human gastric cathepsin E.";
RL Biochem. Biophys. Res. Commun. 168:878-885(1990).
CC -!- FUNCTION: DUE TO ITS INTRACELLULAR LOCATION AND DISTRIBUTION IN
CC LYMPHOID ASSOCIATED TISSUE, IT MAY HAVE A ROLE IN IMMUNE FUNCTION.
CC -!- CATALYTIC ACTIVITY: SIMILAR TO CATHEPSIN D, BUT SLIGHTLY BROADER
CC SPECIFICITY.
CC -!- SUBUNIT: HOMODIMER; LINKED BY A DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M84424; AAA52300.1; -.
DR EMBL; M84413; AAA52300.1; JOINED.
```

```

DR EMBL; M84417; AAA52300.1; JOINED.
DR EMBL; M84418; AAA52300.1; JOINED.
DR EMBL; M84419; AAA52300.1; JOINED.
DR EMBL; M84420; AAA52300.1; JOINED.
DR EMBL; M84421; AAA52300.1; JOINED.
DR EMBL; M84422; AAA52300.1; JOINED.
DR EMBL; J05036; AAA52130.1; -.
DR PIR; A34401; A34401.
DR PIR; A34643; A34643.
DR PIR; A42038; A42038.
DR HSSP; P00794; 3CMS.
DR MEROPS; A01.010; -.
DR MIM; I16890; -.
DR InterPro; IPR001461; -.
DR InterPro; IPR001969; -.
DR Pfam; PF00026; asp; 1.
DR PRINTS; PR00792; PEPsin.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
DR Hydrolase; Aspartyl protease; Glycoprotein; Zymogen; Signal.
FT SIGNAL 1 17
FT PROPEP 18 53
FT CHAIN 54 396
FT MOD_RES 18 18
FT ACT_SITE 96 96
FT ACT_SITE 281 281
FT DISULFID 60 60
FT DISULFID 109 114
FT DISULFID 272 276
FT DISULFID 314 351
FT CARBOHYD 90 90
FT CARBOHYD 220 220
FT CARBOHYD 333 333
SQ SEQUENCE 396 AA; 42793 MW; 40B643C5FB01521E CRC64;

Query Match 11.2%; Score 299.5; DB 1; Length 396;
Best Local Similarity 25.9%; Pred. No. 2.2e-16;
Matches 100; Conservative 65; Mismatches 148; Indels 73; Gaps 16;

QY 50 DPEPEPGRGSEFVEMVNLKSKGQGVVEMTVGSPQPTLNILVDTGSSNFAVGA--- 105
DB 63 DOSAKEP-----LINLVD-----MEYFGTISIGSPQNFVFDTGSSNLWVSVYCT 110

QY 106 APHPFLHRYQRLSTYRDLRKGVVYPTQGWEGELGTLVSPHPGNVTVRANIAAI 165
DB 111 SPACKTHSRFQPSQSTYSQPSQSFQYGTGSLGSLIGADQVSV-EGLTVVGQFQGESV 169

QY 166 TSDKFFINGSNWEGILGLAYAEIARPDSDLPEFFDSLKVQTHVPLNLSLQLCGAGFPLN 225
DB 170 TEPGQTFVD-AEFDGLGLGYPFLA--VGGVTPVFDNMAQ-----NLVDLPMSFYMSN 222

QY 226 QSEVLASVGSMTLIGDHSILVTSGLWYTPIRREWYEVIIIVRVEINGQDLKMDCKEYNY 285
DB 223 PE---GGAGSELIFGYDHSFSGSLNWPVTKQAYQIALDNIQVG--TVMFCE--G 275

QY 286 DKSIVDSGTTNLRPKKVEAAVKSTKAASSTKEPDPGEFLWGEQLVCMQAGTTPWNIPV 345
DB 276 COAIVDTGTLITGPSDKIKQLQNAIGAAP-----VDGEYAVE-----CANLNVMPD 322

QY 346 ISLYLMGEVYNOSFRITLPOQLRPRVEDVATSDQDCKYFALSQSTG----- 393
DB 323 VTFTING-----VPYTLSTPAY--TLDFVDMQFC-----SSGFQGLDIHPAGP 366

QY 394 -TVMGAVIMEGFVVFDRKRGIFA 418
DB 367 LWILGDVFIRQFYSVFDNRGNRVGLA 392

RESULT 10
ID PEPE-CHICK STANDARD; PRT; 383 AA.
AC P16476;
DT 01-AUG-1990 (Rel. 15, Created)

```

```

DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE EMBRYONIC PEPSINOGEN PRECURSOR (EC 3.4.23.-)
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88227903; PubMed=3131317;
RA Hayashi K., Agata K., Mochii M., Yasugi S., Eguchi G., Mizuno T.;
RT "Molecular cloning and the nucleotide sequence of cDNA for embryonic
RT chicken pepsinogen: phylogenetic relationship with prothymosin.";
RL J. Biochem. 103:290-296(1988).
CC -!- DEVELOPMENTAL STAGE: SPECIFICALLY SECRETED DURING THE EMBRYONIC
CC PERIOD IN THE CHICKEN PROVENTRICULUS (GLANDULAR STOMACH)
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D00215; BAA00153.1; -.
DR PIR; A41443; A41443.
DR HSSP; P00794; 3CMS.
DR MEROPS; A01.028; -.
DR InterPro; IPR001461; -.
DR InterPro; IPR001969; -.
DR Pfam; PF00026; asp; 1.
DR PRINTS; PR00792; PEPsin.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
DR Hydrolase; Aspartyl protease; Digestion; Signal; Glycoprotein.
FT SIGNAL 1 16
FT CHAIN 17 383
FT ACT_SITE 94 94
FT ACT_SITE 276 276
FT DISULFID 107 112
FT DISULFID 267 271
FT DISULFID 310 344
FT CARBOHYD 132 132
FT CARBOHYD 204 204
FT CARBOHYD 309 309
FT CARBOHYD 350 350
FT VARIANT 51 51 T->S
SQ SEQUENCE 383 AA; 41719 MW; 1642796871611F54 CRC64;

Query Match 11.2%; Score 298.5; DB 1; Length 383;
Best Local Similarity 25.2%; Pred. No. 2.5e-16;
Matches 90; Conservative 75; Mismatches 125; Indels 67; Gaps 14;

QY 77 YVEMTVGSPQPTLNILVDTGSSNFAVGA---APHPFLHRYQRLSTYRDLRKGVV 132
DB 76 YGTISICTPPQDFTVVFDTGSSNLWVSVCTSPACQSHQMFNPQSSTYKSTQNLIS 135

QY 133 PYTQKWEGLGTLVSPHPGNVTVRANIAAITESDKFFINGSNWEGILGLAYAEIARP 192
DB 136 HYGTGDMEGTVGCDTVTVASLMDTNQLGFLST-SPGQFFVY-VKFDGLGLGYPFLA- 192

QY 193 DDLSEPFDFSLVKQTHV-PNLFSLQLCGAGFPLNQSEVLASVGSMTLIGDHSILVTSGL 251
DB 193 -DGLTPVFDNWNESLLEQNLEFVYLS-----REPMSVMEVGGIDESYFTGSI 240

QY 252 WYTPIRREWYEVIIIVRVEINGQDL--KMDCKEYNDKSIDVSGTTNLRPKKVEAAVK 309
DB 241 NWIPVSYQWQISMDSIIVNKQEIACSSGC-----QAIDTGTSLVAGFASINDIQS 294

```

QY 310 SIKASSTEKFPDGFELGELVLCWQAGTTPNFIFFVISL-----YLMGEVTVNOSFRITILP 365
DB 295 AVGANQNT-----YGEYSV-----NCSHLAMPDVVFVIGGT----- 326
QY 366 QOYLPRVEDVA-----TSDDCYKFAISOSSTGTVMGAVIMEGFYVVEDRARKRIGFA 418
DB 327 -QY--PVPALAYTEONGQGTCTMSSSQNSADWILGDFIRVYYSIFDRANRVGLA 380

RESULT 11
CATD_HUMAN
ID CATD_HUMAN STANDARD: PRT: 412 AA.
AC P07339;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CATHEPSIN D PRECURSOR (EC 3.4.23.5).
GN CTSD.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-85270436; PubMed-3927292;
RA Faust P.L., Kornfeld S., Chirgwin J.M.;
RT "Cloning and sequence analysis of cDNA for human cathepsin D.";
RN Proc. Natl. Acad. Sci. U.S.A. 82:4910-4914(1985).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE-87231068; PubMed-3588310;
RA Westley B.R., May F.E.B.;
RT "Oestrogen regulates cathepsin D mRNA levels in oestrogen responsive
human breast cancer cells";
RN Nucleic Acids, Res. 15:3773-3786(1987).
[3]
RP SEQUENCE FROM N.A.
RX MEDLINE-91299158; PubMed-2069717;
RA Redeker B., Heckendorf B., Grosch H.W., Mersmann G., Hasilik A.;
RT "Molecular organization of the human cathepsin D gene";
RN DNA Cell Biol. 10:423-431(1991).
[4]
RP SEQUENCE OF 1-22 FROM N.A.
RX MEDLINE-94085791; PubMed-8262386;
RA May F.E., Smith D.J., Westley B.R.;
RT "The human cathepsin D-encoding gene is transcribed from an estrogen-
regulated and a constitutive start point.";
RN Gene 134:277-282(1993).
[5]
RP SEQUENCE OF 1-22 FROM N.A.
RX MEDLINE-95021301; PubMed-7935485;
RA Augereau P., Miralles F., Gaudelot C., Parker M.,
Rocheffort H.;
RT "Characterization of the proximal estrogen-responsive element of
human cathepsin D gene";
RN Mol. Endocrinol. 8:693-703(1994).
[6]
RP SEQUENCE OF 170-180.
RC TISSUE=Liver;
RA Hochstrasser D.F., Frutiger S., Paquet N., Bairoch A., Ravier F.,
Pasquali C., Sanchez J.-C., Tissot J.-D., Bjellqvist B., Vargas R.,
Appel R.D., Hughes G.J.;
RL Submitted (JUN-1992) to the SWISS-PROT data bank.
[7]
RP X-RAY CRYSTALLOGRAPHY (3 ANGSTROMS).
RX MEDLINE-93223670; PubMed-8467789;
RA Metcalf P., Fusek M.;
RT "Two crystal structures for cathepsin D: the lysosomal targeting
signal and active site";
RN EMBO J. 12:1293-1302(1993).
[8]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE-93342076; PubMed-8393577;

RA Baldwin E.T., Bhat T.N., Gulnik S., Hosur M.V., Sowder R.C. II,
Cachau R.E., Collins J., Silva A.M., Erickson J.W.;
RT "Crystal structures of native and inhibited forms of human cathepsin
D: implications for lysosomal targeting and drug design.";
RN Proc. Natl. Acad. Sci. U.S.A. 90:6796-6800(1993).
CC -|- FUNCTION: CATHEPSIN D IS AN ACID PROTEASE ACTIVE IN INTRACELLULAR
PROTEIN BREAKDOWN.
CC -|- SUBUNIT: CONSISTS OF A LIGHT CHAIN AND A HEAVY CHAIN.
CC -|- SUBCELLULAR LOCATION: LYSOSOMAL.
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
EMBL; M11233; AAB59529.1; -;
DR EMBL; X05344; CAA28955.1; -;
DR EMBL; M63138; AAA51922.1; -;
DR EMBL; M63134; AAA51922.1; JOINED.
DR EMBL; M63135; AAA51922.1; JOINED.
DR EMBL; M63136; AAA51922.1; JOINED.
DR EMBL; M63137; AAA51922.1; JOINED.
DR EMBL; L12980; AAA16314.1; -;
DR EMBL; S74689; AAD14156.1; -;
DR EMBL; S52557; AAD13868.1; -;
DR PIR; A25771; KHRUD.
DR PDB; 1LYA; 31-JAN-94.
DR PDB; 1LYB; 31-JAN-94.
DR MEROPS; A01.009; -;
DR SWISS-2DPAGE; P07339; HUMAN.
DR MIM; 116840; -;
DR InterPro; IPR001461; -;
DR InterPro; IPR001969; -;
DR Pfam; PF00026; asp; 1.
DR PRINTS; PR00792; PEPsin.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
KW Hydrolase; Aspartyl protease; Glycoprotein; Lysosome; Signal; Zymogen;
3D-structure.
FT SIGNAL 1 18
FT PROPEP 19 64
FT CHAIN 65 412
FT CHAIN 65 161
FT CHAIN 169 412
FT ACT_SITE 97 97
FT ACT_SITE 295 295
FT DISULFID 91 160
FT DISULFID 110 117
FT DISULFID 286 290
FT DISULFID 329 366
FT CARBOHYD 134 134
FT CARBOHYD 263 263
FT STRAND 67 74
FT TURN 75 77
FT STRAND 78 85
FT TURN 86 89
FT STRAND 90 97
FT TURN 98 99
FT STRAND 103 107
FT TURN 108 109
FT TURN 112 113
FT HELIX 115 118
FT TURN 119 119
FT STRAND 123 123
FT HELIX 125 127
FT TURN 129 130
FT STRAND 132 141
FT STRAND 146 158
FT STRAND 172 184
ACTIVATION PEPTIDE.
CATHEPSIN D.
LIGHT CHAIN (PROBABLE).
HEAVY CHAIN (PROBABLE).
N-LINKED (GLCNAC. .).
N-LINKED (GLCNAC. .).

Matches	97	Conservative	63	Mismatches	135	Indels	66	Gaps	13
QY	77	YYEMVTGSPQPTLNILVDTGSSNFVGAAPHPE-----LHRYQQLSSYTRDLRKG	129						
		: :: : :: : : : : : : :							
Db	75	YFGTISGTPQDFTVFDGSSNLWV---PSTYCSSLACLRHFRPDSSTVQGSSET	131						
		: :: : :: : : : : : : :							
QY	130	YVPYTOCKWEGELGTDLVSPHGPNTVTRANIAATESDKFFINGSNWEGILGLAYAEI	189						
		: : : : : : : : : : : : : :							
Db	132	LSITYGTSMTGILGYTVKVGSTEDYNQIFGLSKTSPSLTFLE--APFGILGLAYPSI	189						
		: : : : : : : : : : : : : :							
QY	190	ARPDSDLPEPFDLSLVKOTHV-PNLFSLQLCAGGPPLNQSEVLASVGGSMIIGGDHSLYT	248						
		: : : : : : : : : : : : : :							
Db	190	SSDAT--PVFDNNWNGELYSQDLFSVYLSDD-----EKGSLVFGGIDSSYIT	237						
		: : : : : : : : : : : : : :							
QY	249	GSLWYTPIRREWYEVIIIVRVEINGODLKM--DCKEYNYDKSIVDSGTNLRLEPKKVFEA	306						
		: : : : : : : : : : : :							
Db	238	GSLNWVPVSVEGYQIIMTDSVINGETIACADSC-----QAIVDTGTSLLTGP-----TS	287						
		: : : : : : : : : : : : :							
QY	307	AVKSIIKAASSTKPPDGFWIGLEQLV-CWQAGTTPWNIFFVISLYLMGEVTNQSPRITLP	365						
		: : : : : : : : : : : : :							
Db	288	AISNIQSYIGASK---NLLGENVISCAIDSLPDIVF-----FING	325						
		: : : : : : : : : : : : :							
QY	366	QQYLRPVEDVATGDDCYKFAISSSTGT-----VNGAVIMEGYVVFDRARRKIGFV	419						
		: : : : : : : : : : : : :							
Db	326	IQYPLPASAVILKEDDCTSCLEGMNVDYTGELWILGDDVFIRQYTFVFDNRANQLGLAA	385						
		: : : : : : : : : : : : :							
QY	420	S	420						
		: : : : : : : : : : : : :							
Db	386	A	386						

```

RESULT 13
CATE_CAVPO STANDARD; PRT; 391 AA.
ID CATE_CAVPO AC P25796;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE CATHEPSIN E PRECURSOR (EC 3.4.23.34).
GN CTSE.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RN SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=92355614; PubMed=1644829;
RA Kageyama T., Ichinose M., Tsukada S., Miki K., Kurokawa K., Koiwai O.,
RA Tanji M., Yakabe E., Athauda S.B., Takahashi K.;
RT "Gastric procathepsin E and progastricsin from guinea pig.
RT Purification, molecular cloning of cDNAs, and characterization of
RT enzymatic properties, with special reference to procathepsin E.";
RL J. Biol. Chem. 267:16450-16459(1992).
RN [2]
RN SEQUENCE FROM N.A.
RP RP
RP TISSUE=Gastric mucosa;
RX MEDLINE=96073637; PubMed=8540321;
RA Kageyama T., Ichinose M., Miki K., Moriyama A., Yonezawa S.,
RA Tanji M., Athauda S.B., Takahashi K.;
RT "Isolation, characterization, and structure of procathepsin E and
RT cathepsin E from the gastric mucosa of guinea pig.";
RL Adv. Exp. Med. Biol. 362:211-221(1995).
CC -1- FUNCTION: DUE OT ITS INTRACELLULAR LOCATION AND DISTRIBUTION IN
CC LYMPHOID ASSOCIATED TISSUE, IT MAY HAVE A ROLE IN IMMUNE FUNCTION.
CC -1- CATALYTIC ACTIVITY: SIMILAR TO CATHEPSIN D, BUT SLIGHTLY BROADER
CC SPECIFICITY.
CC -1- SUBUNIT: HOMODIMER; LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use.

```

use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
-----
CC      EMBL; M88653; AAA37052.1; -.
CC      EMBL; S80547; AAB35844.1; -.
CC      PIR; A43356; A43356.
CC      HSSP; P00794; 3CMS.
CC      MEROPS; A01.010; -.
CC      InterPro; IPR0011461; -.
CC      InterPro; IPR001969; -.
CC      Pfam; PF00026; asp; 1.
CC      PRINTS; PR00792; PEPsin.
CC      PROSITE; PS00141; ASP_PROTEASE; 2.
KW      Hydrolase; Aspartyl protease; Glycoprotein; zymogen; Signal.
FT      SIGNAL          1      18
FT      PROPEP         19      53
FT      CHAIN          54      391
FT      ACT_SITE       92      92
FT      ACT_SITE       276     276
FT      DISULFID       56      56
FT      DISULFID      105     110
FT      DISULFID      267     271
FT      DISULFID      309     346
FT      CARBOHYD       86      86
FT      SEQUENCE      391 AA; 42132 MW; 78D216BF8CFCDABD CRC64;
N-LINKED (GLCNAC... ) (POTENTIAL).
      POTENTIAL
      ACTIVATION PEPTIDE (POTENTIAL).
      CATHEPSIN E.
      BY SIMILARITY.
      BY SIMILARITY.
      INTERCHAIN (PROBABLE).
      BY SIMILARITY.
      BY SIMILARITY.
      BY SIMILARITY.

```

Query Match 11.1%; Score 296; DB 1; Length 391;
Best Local Similarity 26.9%; Pred. No. 4e-16;
Matches 98; Conservative 64; Mismatches 130; Indels 72; Gaps 16;

Qy	77	YYVEMTVGSPQPTNLILVDTGSSNFAVGA-----APHFPLHRYQOROLSSSTYRDLRGVYV	132
Db	74	YFGTISIGSSPQNFTVIFDFTGSSNLWVPVSYCTSPACQTHPVFHPSLSSSTYREVGNFSFI	133
Qy	133	PYTGQKWEGLGTDLVSIHPGPNVTVRANIAATSDKPFINGSNWEGILGLAYAEIARP	192
Db	134	QYGTGSVTGLIIGADQVSV-BGLTVVGQFQGESVOEPGKTFVH-AEFDGILGLGYPSLAA-	190
Qy	193	DDSLPEPFDLSLVKQTHVPNLFSLQLCCAGPPLNQSEVLASVGGSMIIGGIDHSLYTGSLW	252
Db	191	-GGVTPYFDNNMAQ-----NLVALPM-----PSVMSSNPGCGSGSELTFGGYDPSHPGSLN	241
Qy	253	YTPIRREWYVEYIIVRVEINGQDLKMDCKEYNYDKSITVSGTNTNLRLPKKVFAAVKSIK	312
Db	242	WVPVTKQAYWQIALDGIQVG--DSVMFCSE--GCQAIVDTGTSLITGP-----PGKIKQLQ	293
Qy	313	AASTEKFPDGFWLGEOLVCWQAGTTPWNITFPVISLYLGEVTVNQSFRI-----TILPOQ	367
Db	294	EALGATVDEGY-----SVQC-----ANLNMMLDVT-----FINGVPYTLNPTA	333
Qy	368	YLRPVEDVATSQDDCYKFAISQSSGTG-----TVMGAVIMEGVYVYVDFDRARX	414
Db	334	Y--TLDFVDGMQVC-----STGEFEGLEIQPAPGLWILGDVFIQFYAVFDRGNRR	393
Qy	415	IGFA 418	
Db	384	VGIA 387	

DEPT. OF THE ARMY

RESULT	ID	PEP4_RABBIT	STANDARD;	PRT;	387 AA.
PEP4_RABBIT	AC	P28713;			
	DT	01-DEC-1992 (Rel. 24, Created)			
	DT	01-DEC-1992 (Rel. 24, Last sequence update)			
	DT	30-MAY-2000 (Rel. 39, Last annotation update)			
	DE	PEPSIN II-4 PRECURSOR (EC 3.4.23.1) (PEPSIN A).			
	OS	Oryctolagus cuniculus (Rabbit).			
	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	OC	Mammalia; Eutheria; Laquomorpha; Leporidae; Oryctolagus			

FT	ACT_SITE	290	290	BY SIMILARITY.
FT	DISULFID	91	160	BY SIMILARITY.
FT	DISULFID	110	117	BY SIMILARITY.
FT	DISULFID	281	285	BY SIMILARITY.
FT	DISULFID	324	361	BY SIMILARITY.
FT	CARBOHYD	134	134	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	258	258	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CONFLICT	15	15	D -> A (IN REF. 2).
FT	CONFLICT	163	163	D -> T (IN REF. 3).
FT	CONFLICT	205	205	K -> N (IN REF. 2).
FT	CONFLICT	262	262	K -> N (IN REF. 2).
SEQ	SEQUENCE	407 AA;	44680 MW;	C423AD4104D95F84 CRC64;

Query Match	11.1%;	Score 295;	DB 1;	Length 407;
Best Local Similarity	26.4%;	Pred. No. 5.1e-16;		
Matches 114;	Conservative 72;	Mismatches 160;	Indels 86;	Gaps 19;

QY	28	IRLPLR-----SGLGGA--PLGLRLPRETDEPEEPGRGRGSFVEMVDNLRGKSGQG 76
DB	22	IRIPLRKFTSIRRTMTVEGGSVEDLILKGPITKYSMQSSPRTKEPVSELLKNYLDQAQ--- 78
QY	77	YVEMTVGSPQPTLNILVDTGSSNFAVGAAPHFL-----HRYYQQLSSTYRDLRKGV 130
DB	79	YGEIGIGTPPQCFTVYFDIGSSNLWVPSIHCKLLDIACWVHHKYSKSDSSTVYKNGTSF 138
QY	131	YVPYTOGKWEGLGTLVSIPIHGPNTVTRANIAAITESDKFF-----INGSNWEG 180
DB	139	DIHYGSGSLGYLSQDTSVSP-----CKSDLGGIRKVEKQIFGEATKQPGVVFIAAKFDG 192
QY	181	ILGLAYAEIARPDSDLEPFDSLVKQTHV--PNLFSIQLCGAGFPLNQSEVLASVGGSMII 239
DB	193	ILGMGYPFIS--VNKVLVPFDNLMKQKLVKKNIFS-----EYLNLR-DPTGQPGGGLML 242
QY	240	GGIDHSLYTGSLWYTPIRREWYVEIIVRVEINGODLKMDCKEYNYDKSIVDSSTNLR 299
DB	243	GGTDSRYHGELSYLNWTRKAYQVHMDQLEV--GSELT--CK--GGCEALVDGTGSLLVG 298
QY	300	PKKVFEAAVKSIIKASSTEFDPDGLGEQLV-CWQAGTTPWNIFPVISLYLMGEVTNQ 358
DB	299	PVDEVKELQKICAVPLIQ-----GEYMIPECVSS-----LPIITFKLGGQ----- 340
QY	359	FRITILPOQYLRPVEDVATSDCCYKFAIS-----QSSTGTVMGAVIMEGFYVVD 409
DB	341	-NYELHPEKYILKVSQAGKT-----ICLSGFMGMDIPPPSGPLWILGDVFIGCYTTFD 393
QY	410	RARKRIGFAVSA 421
DB	394	REYNRVGFAKAA 405

Search completed: September 6, 2001, 16:51:04
Job time: 808 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:49:43 ; Search time 231.42 Seconds
(without alignments)
287.570 Million cell updates/sec

Title: US-09-603-713-3

Perfect score: 2663

Sequence: 1 MASWTGQMGSGMAGVLP.....CLRLRQOHDFADDISLLK 503

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_16.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_unclassified.*
13: sp_vertebrate.*
14: sp_virus.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query % Match	Length	ID	Description
1	2036	76.5	396	4 Q9ULS1	Q9uls1 homo sapien
2	1160.5	43.6	439	4 Q9H2V8	Q9h2v8 homo sapien
3	1155	43.4	514	11 Q9JL18	Q9jl18 mus musculus
4	976	36.7	468	4 Q9NZL2	Q9nzl2 homo sapien
5	971	36.5	396	4 Q9NZL1	Q9nzl1 homo sapien
6	712.5	26.8	213	4 Q9P0D2	Q9p0d2 homo sapien
7	596.5	22.4	255	11 Q9R1P7	Q9rip7 mus musculus
8	332.5	12.5	391	5 Q9VX7	Q9vxp7 boophilus m
9	332	12.5	354	5 Q9GMY8	Q9gmy8 boophilus m
10	309	11.6	384	13 Q9DEC2	Q9dec2 xenopus lae
11	308	11.6	385	13 Q9DEC4	Q9dec4 rana catesb
12	307	11.5	346	6 Q9TS27	Q9ts27 bos taurus
13	307	11.5	372	5 Q9VLK3	Q9vlk3 drosophila
14	305	11.5	386	6 Q9GMV7	Q9gmv7 rhinolophus
15	305	11.5	387	6 Q9GMV8	Q9gmv8 sorex ungu
16	304.5	11.4	383	13 Q9DEC3	Q9dec3 xenopus lae
17	304.5	11.4	387	13 Q9DDV5	Q9ddv5 salvelinus
18	304	11.4	387	6 Q9GMV9	Q9gmv9 suncus mur
19	303.5	11.4	383	13 Q9DE45	Q9de45 salvelinus

20	302.5	11.4	376	13 Q9PUR8	Q9pur8 pleuronecte
21	302	11.3	384	13 Q91322	Q91322 rana catesb
22	301	11.3	382	13 Q9PRG9	Q9prg9 gallus gall
23	301	11.3	423	5 Q9VKP7	Q9vvp7 drosophila
24	296.5	11.1	396	13 Q93428	Q93428 chionodraco
25	295.5	11.1	386	6 Q9CMV6	Q9cmv6 canis famil
26	293	11.0	399	13 Q93458	Q93458 podarcis sl
27	290.5	10.9	381	6 Q9GK11	Q9gk11 camelus dro
28	289.5	10.9	399	13 Q9DD89	Q9dd89 brachydanio
29	287.5	10.8	444	5 Q21966	Q21966 caenorhabdi
30	285.5	10.7	380	6 Q28950	Q28950 sus scrofa
31	284	10.7	398	13 P87370	P87370 oncorhynch
32	284	10.7	427	5 P91802	P91802 schistosoma
33	283.5	10.6	378	13 Q9PUR9	Q9pur9 pleuronecte
34	283	10.6	422	5 Q9G906	Q9g906 onchocerca
35	281	10.6	389	13 Q9PWK1	Q9pwk1 gallus gall
36	280.5	10.5	396	13 Q9DEX3	Q9dex3 clupea hare
37	279.5	10.5	390	6 Q9CK10	Q9ck10 camelus dro
38	278	10.4	370	6 Q9TTW1	Q9ttw1 bos taurus
39	278	10.4	387	6 Q9N2D4	Q9n2d4 callithrix
40	278	10.4	389	13 Q9W643	Q9w643 gallus gall
41	277	10.4	446	5 Q9N9H3	Q9n9h3 necator ame
42	275	10.3	385	6 Q29080	Q29080 sus scrofa
43	275	10.3	413	3 Q14413	Q14413 pichia angu
44	273.5	10.3	410	5 Q9VQ13	Q9vq13 drosophila
45	273	10.3	388	6 Q9GMY2	Q9gmy2 oryctolagus

ALIGNMENTS

RESULT 1

Q9ULS1 PRELIMINARY; PRT; 396 AA.
ID Q9ULS1
AC Q9ULS1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE KIAA1149 PROTEIN (FRAGMENT).
GN KIAA1149.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=20039618; PubMed=10574461;
RA Hirose M., Nagase T., Ishikawa K., Kikuno R., Nomura N., Ohara O.;
RT "Characterization of cDNA clones selected by the Genemark analysis
RL DNA Res. 6:329-336(1999).
DR EMBL; AB032975; BAA86463.1; -.
DR HSSP; P56272; 1AM5.
DR InterPro; IPR001461; -.
DR Pfam; PF00026; asp; 2.
DR PRINTS; PR00792; PEPSIN.
FT NON_TER 1 1
SQ SEQUENCE 396 AA; 44428 MW; A2CBCD52DCC089E0 CRC64;

Query Match 76.5%; Score 2036; DB 4; Length 396;
Best Local Similarity 100.0%; Pred. No. 4.7e-156;
Matches 384; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	120	SSYRLRLRGVYVPYQGWEGELGTDLVSIHPGNVTVRANAAITSDKFFINGSNW	179
DB	13	SSYRLRLRGVYVPYQGWEGELGTDLVSIHPGNVTVRANAAITSDKFFINGSNW	72
QY	180	GILGLAYAEIARDDSLVQTHVPNLFSLQCGAGFPLNQSEVLASVGGSMII	239
DB	73	GILGLAYAEIARDDSLVQTHVPNLFSLQCGAGFPLNQSEVLASVGGSMII	132

```
QY 240 GGIDHSLYTGSLWYTPIRREWEYEVIIVRVEINGODLKMCKEYNDKSIYDSGTTNRL 299
|||||
Db 133 GGIDHSLYTGSLWYTPIRREWEYEVIIVRVEINGODLKMCKEYNDKSIYDSGTTNRL 192
|||||
QY 300 PKKVFEEAAVKSIIKAASSTKFPDGEWLGQVLCVQAGTTPWNIFFVLSLYLMGEVTSQF 359
|||||
Db 193 PKKVFEEAAVKSIIKAASSTKFPDGEWLGQVLCVQAGTTPWNIFFVLSLYLMGEVTSQF 252
|||||
QY 360 RITILPQOYLRVEDVATVSDQCYFAISQSSTGTVMGAVIMEGYVVFDRARKRIGFAV 419
|||||
Db 253 RITILPQOYLRVEDVATVSDQCYFAISQSSTGTVMGAVIMEGYVVFDRARKRIGFAV 312
|||||
QY 420 SACHVDEPRTAAVEGPFVTLDMDCGYNIPQTDSTLMTIAYVMAAICALFMLPLCLMW 479
|||||
Db 313 SACHVDEPRTAAVEGPFVTLDMDCGYNIPQTDSTLMTIAYVMAAICALFMLPLCLMW 372
|||||
QY 480 CQWRCLRLCRLQHQDDFADDISLLK 503
|||||
Db 373 CQWRCLRLCRLQHQDDFADDISLLK 396
|||||
RESULT 2
Q9H2V8 PRELIMINARY; PRT; 439 AA.
AC Q9H2V8
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE CDAL3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PHOCHROMOCYTOMA;
RA Li Y., Huang Q., Peng, Y., Song H., Yu Y., Xu S., Ren S., Chen Z.,
RA Han Z.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF21252; AAG41783.1; -.
SQ SEQUENCE 439 AA; 48275 MW; 02EC0E0E50F11602 CRC64;
```

```
Query Match 43.6%; Score 1160.5; DB 4; Length 439;
Best Local Similarity 50.1%; Pred. No. 2.3e-85;
Matches 220; Conservative 78; Mismatches 134; Indels 7; Gaps 4;

QY 65 MVDNLRKSGQGYVEMTSGPPQTLNLLVDTGSSNFVAGAAPHPFLHRYQRLSSTYR 124
|||||
Db 1 MVDNLRKSGQGYVEMTSGPPQTLNLLVDTGSSNFVAGAAPHPFLHRYQRLSSTYR 60
|||||
QY 125 DLKRGYVPTQCKWEGELGTDLVSTPHGPNVTVRANIAALTESDKFFINGSNWEGILGL 184
|||||
Db 61 SKGFDVTYKVTQSGWTFVGEDLVTPKGFNTSFLVNIATIFESNFFLPGLKNGILGL 120
|||||
QY 185 AYAEIARPDSDLFPFFDSLVKQTHVPNLFSLQLCGAGFPINQSEVLASVGGSMIIGDIDH 244
|||||
Db 121 AYATLAKPSSLETFFDSLVTKQNIPIVFSMQCGAGLPVAGS--GTNGGSLVLGGIEP 177
|||||
QY 245 SLYTGSLWYTPIRREWEYEVIIVRVEINGODLKMCKEYNDKSIYDSGTTNRLPKKVF 304
|||||
Db 178 SLYKGDIIWYTPKKEWYQIEILKLGQSLNLCREYNADKAIYDSGTTNRLPKKVF 237
|||||
QY 305 EAAVKSIKAASSTKFPDGEWLGQVLCVQAGTTPWNIFFVLSLYLMGEVTSQF 364
|||||
Db 238 DAVVEAVARASLPIFSDGFWTQSLACWTNSETPHWYFKIYLRDENSSFRITIL 297
|||||
QY 365 PQOYLRVEDVATVSDQCYFAISQSSTGTVMGAVIMEGYVVFDRARKRIGFAVSACHV 424
|||||
Db 298 POLYIQPMGAGLNY-ECYRFGISPSNVALVIGATMEGYVVFDRARKRIGFAVSACHV 356
|||||
QY 425 HDEPRTAAVEGPFVTLDMDCGYNIPQTDSTLMTIAYVMAAICALFMLPLCLMW 483
|||||
```

```
Db 357 IAGAAVSEISGPFSTEDVASNCVPAQSLSEPIILWIVSYALMSVCGAILLVLLVLLPFR 416
QY 484 CLRCRLQHQDDFADDISLL 502
|||||
Db 417 CQR--RPRDPEVNDSSL 433
|||||
RESULT 3
Q9JL18 PRELIMINARY; PRT; 514 AA.
AC Q9JL18
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE ASPARTYL PROTEASE 1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Choi D.K., Sugano S., Sakaki Y.;
RT "Molecular characterization of the mouse Aspl gene, a homolog of the
RL human Aspl (Down Syndrome Region aspartyl protease).";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF216310; AAF36599.1; -.
DR InterPro; IPR001461; -.
DR InterPro; IPR001969; -.
DR Pfam; PF00026; asp; 3.
DR PRINTS; PR00792; PEPsin.
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_2.
KW Protease.
SQ SEQUENCE 514 AA; 55799 MW; A70725F2C1DF5B47 CRC64;
```

```
Query Match 43.4%; Score 1155; DB 11; Length 514;
Best Local Similarity 48.5%; Pred. No. 8e-85;
Matches 225; Conservative 76; Mismatches 143; Indels 20; Gaps 5;

QY 16 AGVLPAGHTQHGIRLPSGLGGAPLRLPRETDEEPEEGRGSGFVEMVDNLRKSGQ 75
|||||
Db 43 ASAVPLGLTP---ELPRADGLA---LALEPVRA-----ANFLAMVDNLQDSSR 86
|||||
QY 76 GYVEMTSGPPQTLNLLVDTGSSNFVAGAAPHPFLHRYQRLSSTYRDLRKGVVYPT 135
|||||
Db 87 GYVLEMLGTTPKQVQIILVDTGSSNFVAGAPHSYIDTYFDSSESTYHSKGFDTVKYT 146
|||||
QY 136 QCKWEGELGTDLVSTPHGPNVTVRANIAALTESDKFFINGSNWEGILGLAYAEIARPD 195
|||||
Db 147 QGSWTFVGEDLVTPKGFNTSFLVNIATIFESNFFLPGLKNGILGLAYAAKAPSS 206
|||||
QY 196 LEPPFDSLVKQTHVPNLFSLQLCGAGFPINQSEVLASVGGSMIIGDIDHSLYTGSLWYTP 255
|||||
Db 207 LETFFDSLVQAQAKIPDIFSMQCGAGLPVAGS--GTNGGSLVLGGIEPSLYKGDIIWTP 263
|||||
QY 256 IREWEYEVIIVRVEINGODLKMCKEYNDKSIYDSGTTNRLPKKVFEEAAVKSIIKAAS 315
|||||
Db 264 IKEEWYQIEILKLGQSLNLCREYNADKAIYDSGTTNRLPKKVFEEAAVKSIIKAAS 323
|||||
QY 316 STEKPDGFWLGQVLCVQAGTTPWNIFFVLSLYLMGEVTSQF 375
|||||
Db 324 LIPEFSDGFWTQSLACWTNSETPHWYFKIYLRDENASFRITILPOLYIQPMGA 383
|||||
QY 376 ATSODCYKFAISQSSTGTVMGAVIMEGYVVFDRARKRIGFAVSACHVDEPRTAAVEG 435
|||||
Db 384 GFNY-ECYRFGISSTNVALVIGATMEGYVVFDRARKRIGFAVSACHVDEPRTAAVEG 442
|||||
QY 436 PFVTLDMDCGYNIPQTDSTLMTIAYVMAAICALFMLPLCLMW 479
|||||
Db 443 PFSTEDIASNCVPAQSLSEPIILWIVSYALMSVCGAILLVLLVLL 486
|||||
RESULT 4
```

```
Q9NZL2
ID Q9NZL2 PRELIMINARY; PRT; 468 AA.
AC Q9NZL2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE ASPARYL PROTEASE.
GN BACE2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Solans A., Estivill X., de la Luna S.;
RT "Cloning of a novel mammalian aspartyl protease.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF168276; AAF35835.1; -.
DR InterPro; IPR001461; -.
DR Pfam; PF00026; asp. 3.
DR PRINTS; PR00792; PEPsin.
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_2.
KW Protease.
SQ SEQUENCE 468 AA; 50324 MW; 717E0920126A0142 CRC64;

Query Match 36.7%; Score 976; DB 4; Length 468;
Best Local Similarity 41.1%; Pred. No. 1.9e-70;
Matches 199; Conservative 75; Mismatches 136; Indels 74; Gaps 6;

QY 20 PAHQGHGIRLPLRSLGGLGAPLGLRLPRETDEPEEGRGRSFVEMVNLKSGQGYV 79
DB 52 PAERHADGLALEPALA-----SPAGANFLAMVDNLQDGRGYL 94

QY 80 EMTVSPPTNLIVDTGSSNFVAGAAHPFLHRYQRLSTYRDLKRGVYVPTQGW 139
DB 95 EMLGTPKQLQILVDTGSSNFVAGTSHSYIDYFDTERTSSTYRSGFDVTVKYTQGSW 154

QY 140 EGEGLTDLVSIPIGPNVTVRANIAAITEDEKFFINGSNWEGILGLAYAIARPDLSLEPF 199
DB 155 TGFVGEDLVTPKGTSTFLVNIATIFESNFFLPKIKWNGILGLAYATLAKPSSLETFF 214

QY 200 FDSLKQTHVPLFSLQCGAGFPLNQSEVLASVGGSMIIGDHSHTYGLSWYTPIRRE 259
DB 215 FDSLVTQANIPNVFSQMCGAGLPVAGS---GTNGGSLVLGGIEPSLYKGDWYTPIKEE 271

QY 260 WYVEIIVRVEINGODLKMCKEYNDKSIYDSGTNNLRPKKVEAAVKSIAASSTEK 319
DB 272 WYQIEILKLEIGGSLNLDREYNADRAIVDSGTTLRLPKQKVEDAVVEAVARASLL-- 329

QY 320 FPDGFWLGEQLVCWQAGTPWNIFPVISLYLMGEVNTQSFRTITLPQOYLRPVEDVATSQ 379
DB 330 -----YIQPMWAGALNY 341

QY 380 DDCYKFAISQSTGTVMGAVIMEGVYVFDARKRIGFAVSACHVDFEFTAAGEGPFVT 439
DB 342 -ECYREGISPSTNALVIGATVMEGVYVFDRAQKRVGAFAAEIAGAAVSEISGPEST 400

QY 440 LDMEDCGNIPOTDESTLTMTIAYVMAATC-ALFMLEPLCLMVGCOWRCLRCRQHQHDFADD 498
DB 401 EDVASCNCPAQSLSEPIWIVSYALMSVCGAIIILLVLLLPFRQCR--RPRDPEVVDND 458

QY 499 ISLL 502
DB 459 ESSL 462

RESULT 5
Q9NZL1
ID Q9NZL1 PRELIMINARY; PRT; 396 AA.
AC Q9NZL1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
```

```
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE ASPARYL PROTEASE.
GN BACE2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Solans A., Estivill X., de la Luna S.;
RT "Cloning of a novel mammalian aspartyl protease.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF168277; AAF35836.1; -.
DR InterPro; IPR001461; -.
DR Pfam; PF00026; asp. 2.
DR PRINTS; PR00792; PEPsin.
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_2.
KW Protease.
SQ SEQUENCE 396 AA; 43013 MW; 5023A7AF391CEAC9 CRC64;

Query Match 36.5%; Score 971; DB 4; Length 396;
Best Local Similarity 50.9%; Pred. No. 3.7e-70;
Matches 189; Conservative 55; Mismatches 97; Indels 30; Gaps 5;

QY 20 PAHQGHGIRLPLRSLGGLGAPLGLRLPRETDEPEEGRGRSFVEMVNLKSGQGYV 79
DB 52 PAERHADGLALEPALA-----SPAGANFLAMVDNLQDGRGYL 94

QY 80 EMTVSPPTNLIVDTGSSNFVAGAAHPFLHRYQRLSTYRDLKRGVYVPTQGW 139
DB 95 EMLGTPKQLQILVDTGSSNFVAGTSHSYIDYFDTERTSSTYRSGFDVTVKYTQGSW 154

QY 140 EGEGLTDLVSIPIGPNVTVRANIAAITEDEKFFINGSNWEGILGLAYAIARPDLSLEPF 199
DB 155 TGFVGEDLVTPKGTSTFLVNIATIFESNFFLPKIKWNGILGLAYATLAKPSSLETFF 214

QY 200 FDSLKQTHVPLFSLQCGAGFPLNQSEVLASVGGSMIIGDHSHTYGLSWYTPIRRE 259
DB 215 FDSLVTQANIPNVFSQMCGAGLPVAGS---GTNGGSLVLGGIEPSLYKGDWYTPIKEE 271

QY 260 WYVEIIVRVEINGODLKMCKEYNDKSIYDSGTNNLRPKKVEAAVKSIAASSTEK 319
DB 272 WYQIEILKLEIGGSLNLDREYNADRAIVDSGTTLRLPKQKVEDAVVEAVARASLL-- 331

QY 320 FPDGFWLGEQLVCWQAGTPWNIFPVISLYLMGEVNTQSFRTITLPQOYLRPVEDVATSQ 379
DB 332 FSDGFWTGSQACWNTSETPNWYFPKISYLRDENSRSFRITILPQK-LRVLQ----- 384

QY 380 DDCYKF-AISQ 389
DB 385 --CLAFPLSQ 393

RESULT 6
Q9PD02
ID Q9PD02 PRELIMINARY; PRT; 213 AA.
AC Q9PD02;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HSPC104 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=CORD BLOOD;
RA Zhang Q.H., Ye M., Zhou J., Shen Y., Wu X.Y., Guan Z.Q., Wang L.,
RA Fan H.Y., Mao Y.F., Dai M., Huang Q.H., Chen S.J., Chen Z.;
```

RT	"Human partial CDS cloned from cd34+ stem cells.";	
RL	Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AF161367; AAF28927.1; -	
DR	InterPro; IPR001461; -	
DR	Pfam; PF00026; asp; 1.	
FT	NON_TER	1
SO	SEQUENCE	213 AA; 24338 MW; EC9D3FA31CFA835C CRC64;
Query Match		
Best Local Similarity		
Matches 137; Conservative		
26.8%; Score 712.5; DB 4; Length 213;		
83.5%; Pred. No. 1e-49;		
Mismatches 12; Indels 11; Gaps 1;		
OY	240 GGIDHSLTGSLWTPTRREWEYEVIIVRVEINGDLKMDCKEYNDKSIDVSGTTNLR	299
DB	1 GGIDHSLTGSLWTPTRREWEYEVIIVRVEINGDLKMDCKEYNDKSIDVSGTTNLR	60
OY	300 PKKVEAAVKSIKAASSTKPKDGFGLGQVLCVQAGTTPWNIFFVISLYLMGEVTNQSF	359
DB	61 PKKVEAAVKSIKAASSTKPKDGFGLGQVLCVQAGTTPWNIFFVISLYLMGEVTNQSF	120
OY	360 RTILPQOYLRLPVEDVATSDQDCYKFAISQSGSTGTVMGAVIMEG	403
DB	121 RTILPQOYLRLP-----WKMPRPKTTVTVCHLTVIHG	153
RESULT	7	
Q9R1P7	PRELIMINARY;	PRT; 255 AA.
AC	Q9R1P7;	
DT	01-MAY-2000 (Tremblrel. 13, Created)	
DT	01-MAY-2000 (Tremblrel. 13, Last sequence update)	
DT	01-JUN-2000 (Tremblrel. 14, Last annotation update)	
DE	ASPARTYL PROTEASE (FRAGMENT).	
OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
OX	NCBI_TaxID=10090;	
RN	SEQUENCE FROM N.A.	
RA	Accarino M., Fumagalli P., Taramelli R., Ottolenghi S.;	
RT	"Cloning of a gene from chromosome 21 Down Region encoding a potential	
RT	transmembrane protease."	
RL	Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AF051150; AAD45964.1; -	
DR	InterPro; IPR001461; -	
DR	InterPro; IPR001969; -	
DR	Pfam; PF00026; asp; 2.	
DR	PRINTS; PR00792; PEPsin.	
DR	PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.	
KW	Protease.	
FT	NON_TER	1
SO	SEQUENCE	255 AA; 28685 MW; 53DE317815996D63 CRC64;
Query Match		
Best Local Similarity		
Matches 109; Conservative		
22.4%; Score 596.5; DB 11; Length 255;		
47.8%; Pred. No. 3e-40;		
Mismatches 44; Indels 1; Gaps 1;		
OY	252 WYTPIRREWEYEVIIVRVEINGDLKMDCKEYNDKSIDVSGTTNLRPKKVEAAVKSI	311
DB	1 WYTPIRREWEYEVIIVRVEINGDLKMDCKEYNDKSIDVSGTTNLRPKKVEAAVKSI	60
OY	312 KAASSTKPKDGFGLGQVLCVQAGTTPWNIFFVISLYLMGEVTNQSFRTILPQOYLRLP	371
DB	61 ARTSLPESDGFQWTAQLACNTSETWPAYPKISILRDENASRFRITILPOLYIQP	120
OY	372 VEDVATSDQDCYKFAISQSGSTGTVMGAVIMEGYVYVDFDRKRIGFVAVSACHVDFEPTA	431
DB	121 MMGAGFNY-ECYRFGISSSTNALVIGATVMEGYVYVDFDRKRIGFVAVSACHVDFEPTA	179
OY	432 AVSGPFVTLDMEDCGYNIPQTDSTLMTATAYVMAICALFMLPLCLMW	479

Db	180 EISGPFSTEDIASNCVPAQALNEPILWIVSYALMSVCGAILLVILL 227
RESULT	8
Q9VKP6	PRELIMINARY; PRT; 391 AA.
ID	Q9VKP6;
AC	Q9VKP6;
DT	01-MAY-2000 (Tremblrel. 13, Created)
DT	01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT	01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE	CG17134 PROTEIN.
GN	CG17134.
OS	Drosophila melanogaster (Fruit fly).
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC	Ephydroidea; Drosophilidae; Drosophila.
OX	NCBI_TaxID=7227;
RN	[1]
RA	SEQUENCE FROM N.A.
RA	STRAIN=BERKELEY;
RA	MEDLINE=20196006; PubMed=10731132;
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA	Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA	Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA	April J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhattacharya D., Bolshakov S.,
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann
RA	Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA	Liu X., Mattet B., McIntosh T.C., McLeod M.P., McPherson D.,
RA	Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA	Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.,
RA	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA	Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA	Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA	Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA	Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA	Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith
RT	"The genome sequence of Drosophila melanogaster."
RL	Science 287:2185-2195(2000).
CC	-!- SIMILARITY: THE PROTEASE BELONGS TO THE PEPTIDASE FAMILY A1;
CC	KNOWN AS THE EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC	EMBL; AE003630; AAF53016.1; -
DR	HSSP; P00794; 4CMS.
DR	FlyBase; FBgn0032304; CG17134.
DR	InterPro; IPR001461; -
DR	InterPro; IPR001969; -
DR	Pfam; PF00026; asp; 1.
DR	PRINTS; PR00792; PEPsin.
DR	PROSITE; PS00141; ASP_PROTEASE; 2.
KW	Aspartyl protease; Hydrolase.
SO	SEQUENCE 391 AA; 42016 MW; 99A18E7131025E1B CRC64;

Query Match 12.5%; Score 332.5; DB 5; Length 391;
Best Local Similarity 29.1%; Pred. No. 1.2e-18;
Matches 105; Conservative 61; Mismatches 146; Indels 49; Gaps 13;

QY 67 DNLGRSGGYYVEMTVGSPPTNLIVDTGSSNFAVGAAPHF-----LHRYQRLS 120
DB 66 ENLHSMNNEYGVIAIGTPEQRNLFDTGSANLWPSASCPASTACORHNKYDSSAS 125
QY 121 STYRDLRKGVYVPYTOCKWEGELGTDLVSIHPGPNVTVRANIAATESDKFINGSNWEG 180
DB 126 STYVANGEEFAIEYGGSLGSLNSDIVTIA-GISIQNTGFEALSEPCTTFVD-APFAG 183
QY 181 ILGLAYAEIARPDSDLEPFDSLVKTHVPN-LFSLQLCAGAPPLNOSSEVLASVGGSMII 239
DB 184 ILGLAFSAIA--VDGVTPTFDNMISQGLDEPVISFLKRGQ-----TAVRGELIL 233
QY 240 GGIDHSLTGLSLWYTPIRREWEYEVIIIVRVEINGODLMDCKEYNDKSDVSDSGTTLRL 299
DB 234 GGIDSSLYRGLTYVPVPAVQFVKVNTIKTNGTLNCGC-----QALADTGTSLIAV 287
QY 300 PKRVFAAVKSIKAASSTEKEPDGFWLGEOLV-CWQAGTTPWNIFPVISLYLMGEVYNOS 358
DB 288 PLAAYRKINRLQATDND-----GEAFVRCGRVSS-----LPKVLNINIGTV- 329
QY 359 FRITILPOQYLRPVEDVATSDDCYK-FAISQSSTGTVMGAVIMEGFYVYVDRARRIGF 417
DB 330 --FTLAPROYI--VKVTONGQTYCMSAFTYMEGLSFWILGDFVIGKFTYVFDKGNRIGF 385
QY 418 A 418
DB 386 A 386

RESULT 9
Q9GYX7 PRELIMINARY; PRT; 354 AA.
AC Q9GYX7;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DE HEME-BINDING ASPARTIC PROTEINASE (FRAGMENT).
OS Boophilus microplus (Cattle tick).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Parasitiformes; Ixodida; Ixodidae; Boophilus.
OX NCBI_TaxID=6941;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PORTO ALEGRE; TISSUE=OVARY;
RA Sorgine M.H.F., Logullo C., Zingali R.B., Paiva-Silva G.O.,
RA Juliano L., Oliveira P.L.;
RT "A heme-binding aspartic proteinase from the eggs of the hard tick
Boophilus microplus".
RL J. Biol. Chem. 0:0-0(2000).
DR EMBL; AF286865; AAG00993.1;
FT NON_TER 1
SQ SEQUENCE 354 AA; 38512 MW; B50D7C638CF27091 CRC64;

Query Match 12.5%; Score 332; DB 5; Length 354;
Best Local Similarity 26.2%; Pred. No. 1.1e-18;
Matches 101; Conservative 74; Mismatches 135; Indels 76; Gaps 15;

QY 66 VDLNRGK-----SGOGYVEMTVGSPPTNLIVDTGSSNFAVGAAPHF- 111
DB 13 VTEIRGALGDPIPIILNTNNMQFYGIITGTPQSFKLLMDTSGSNFWV-----PSIN 67
QY 112 -----HRYQRLSSTYRDLRGVYVPYTOCKWEGELGTDLVSIHPGPNVTVRANI 162
DB 68 CDQSMACROHAKYDSSKSTFTSGRYIRIRYSGGVVGRGITSIDNVGV--GPATVYQYKF 125
QY 163 AATESDKFINGSNWEGILGLAYAEIARPDSDLEPFDSLVKTHVPN-LFSLQLCGAG 221
DB 126 AEMDSDGKLFNRNAYKDGIPLGAPPSISQ--NNQLPLFDANVKGQVYVQAVFSLYL--SK 181

QY 222 FPLNQSEVLASVGGSMIIIGIDHSLTGLSLWYTPIRREWEYEVIIIVRVEINGODLKM-DC 280
DB 182 QPSEQN-----GGEYFGGINAQRVTAIHYVPVQAHAHQVVMNDININVQGTCLCVGGC 235
QY 281 KEINYDKSIVSDSTTNLRPKKVFEEAAVKSIAASSTEKEPDPGFWLGEOLVCHQAGTTPW 340
DB 236 -----PTVVDSSTLSLGP-----SADVETLNRVIGATKTAAGY-----FEVNCATI 277
QY 341 NFPVVISLVLMBG---VTNQSFRITILPOQYLRPVEDVATSDDCYKFAISQSSTGT--- 394
DB 278 SSLPPTITFNLNGKSPFLOGEAYTIRI-----PLTTGGEOCTRISDESASGTNLW 327
QY 395 VMGAVIMEGFYVYVDRARRKRIGFAVS 420
DB 328 ILGAVFTQTYTVTFDRAQNVRGFA 353

RESULT 10
Q9DEC2 PRELIMINARY; PRT; 384 AA.
AC Q9DEC2;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DE PEPsinogen A.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Ikuzawa M., Inokuchi T., Kobayashi K., Yasumasu S.;
RT "Molecular Cloning of Pepsinogens in Adult Xenopus laevis and Bullfrog
Rana catesbeiana".
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB045380; BAB20798.1;
SQ SEQUENCE 384 AA; 41888 MW; 729DD2E7953D9072 CRC64;

Query Match 11.6%; Score 309; DB 13; Length 384;
Best Local Similarity 26.7%; Pred. No. 8.9e-17;
Matches 112; Conservative 66; Mismatches 159; Indels 82; Gaps 20;

QY 28 IRLPL3SLGG---APLGLRLPRETDEEPEPCRR-----GSFVEMVNLKSGQGY 77
DB 17 VKYPL3KGESFRNRLQRLGL-LGDLKKNYPNPASKYFPTLAQSSAETLQNYNDIE---Y 72
QY 78 YVEMTVGSPPTNLIVDTGSSNFAVGAAPHF-----HRYQRLSSTYRDLRGV 130
DB 73 YGTISIGTPPEFTVIFDTGSANLW---PSVYCSSQACSNHNRFPQOSTFQATNTPV 129
QY 131 YVPYTOCKWEGELGTDLVSIHPGPNVTVRANIAATESDK-EFTINGSNWEGILGLAYAEI 189
DB 130 SIQYTGSGMSGFLGYDTLQV---GNIQISNQMGLESEPCSFLYYSPFDGILGLAPPSI 186
QY 190 ARPDSDLEPFDSLVKTHVP-NLFSLQLCAGAPPLNOSSEVLASVGGSMIIIGIDHSLYT 248
DB 187 A--SSQATPVFDNMWSOGLIPQNLFSVYLSDDG-----QTGSYVLFVGGVDNSYYS 234
QY 249 GSLWYTPIRREWEYEVIIIVRVEINGQDL--KMDCKEYNDKSDVSDGTTLNRPLPKKVFEEA 306
DB 235 GSLNWVPLTAETVWQITLDSVINGOVIAQSQC-----QAIVDTGSLMTGPSTPI-A 287
QY 307 AVKSIAASSTEKEPDPGFWLGEOLV-CWQAGTTPWNIFPVISLYLMGEVYNOSFRITLP 365
DB 288 NIQNY--GASODSN-----GOYVINCNNISNMPITV-----TING 322
QY 366 QOY-LRPVEDVATSDDCYK-FAISQSSTGT---VMGAVIMEGFYVYVDRARRKRIGFA 418
DB 323 VQYPLPSAVVRNQOCCSGFOAMNLPNTNSGDLWLGLDFVIRQYFTVFDRAVNYAIA 381

RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.-H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AB003621; AAF52686.1; -.
DR HSSP: P00797; 2REN.
DR FlyBase: FBgn0032049; CG13095.
DR InterPro: IPR001461; -.
DR InterPro: IPR001969; -.
DR Pfam: PF00026; asp; 1.
DR PRINTS: PR00792; PEPIN.
DR PROSITE: PS00141; ASP_PROTEASE; UNKNOWN_2.
SQ SEQUENCE 372 AA; 40080 MW; D45469E8AD72FCCE CRC64;

Query Match 11.5%; Score 307; DB 5; Length 372;
Best Local Similarity 26.0%; Pred. No. 1.2e-16;
Matches 101; Conservative 55; Mismatches 151; Indels 82; Gaps 12;

QY 44 RLP--RETDEEPEEPCRGSFVEMVNDLKGKSGQGYVEMTVGSPQTLNILDVTCSSNF 101
DB 49 QLPFLSRVDEE-----QLSNSMNMAYIGAISIGTPAQSEKVLFDGSSNL 93
QY 102 AVCA----APHPLHRYQRLSSTYRDLRKGVVVPYTOCKWEGELGTLVSLPHGPNVT 157
DB 94 WVPSTNCKSDACLTHNQYDSSASTTVANGESFSIQYGTGSLTGYLSTDTVDV-NGLSIQ 152
QY 158 VRANIAAITESDKFFNGSNWEGILGLAYAEIARPDLSLEPPFDSLVKQTHVPN-LFSLQ 216
DB 153 SQTFEAESTNEPGTNF-NDANFDGILCMAYESLA--VDGVAPPYNNMVSGQLVDSNVSEFY 209
QY 217 LCAGAPLNQSEVLASVSGSMIGGIDHLSLYTGLSNTYPIRREWEYVEIIVRVEINGQDL 276
DB 210 LARDG-----TSTMGGELIPGSDASLYSGALTYVPISEQYQWFTMAGSIDGYSL 261
QY 277 KMDCKEYNDKIVDSGTTNLRLP-----KKVEAAVKSIKAASSTKFPDGFWLGEQ 329
DB 262 CDC-----QATADTGTSLIVAPYNAVYTLSELNVGDEGYLDCSVSLPD----- 308
QY 330 LVCWQAGTTPWNIPFVVISLVMGEVNTNQSFRTITLPQQYLRPVEDVATSDCCYKFAISQ 389
DB 309 -VTFNIGGTNF-----VLKPSAYIIQSDGNCMSAFEYM 340
QY 390 SSTGTWAGVIMGFGYVVFDRARKRIGFA 418
DB 341 GTDFWILGDVFIQGYTYEFDLGNRIGFA 369

RESULT 14
Q9GMV7

ID Q9GMV7 PRELIMINARY; PRT; 386 AA.
AC Q9GMV7;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE PEPINOGEN A.
GN PGNA.
OS Rhinolophus ferrumequinum.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Chiroptera; Microchiroptera; Rhinolophidae;
OC Rhinolophus.
OX NCBI_TaxID=59479;
RN [1]
RA Narita Y., Oda S., Takenaka O., Kageyama T.;
RT "Phylogenetic position of Insectivora inferred from the cDNA sequences
of pepsinogen A and C.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB047245; BAB11751.1; -.
SQ SEQUENCE 386 AA; 41591 MW; 917EE04D3166C3A4 CRC64;

Query Match 11.5%; Score 305; DB 6; Length 386;
Best Local Similarity 29.1%; Pred. No. 1.9e-16;
Matches 104; Conservative 60; Mismatches 130; Indels 64; Gaps 16;

QY 77 YVEMTVGSPQTLNILDVTCSSNFVAVG-----AAPHLHRYQRLSSTYRDLRKGVV 132
DB 74 YFGTIGITPPQEFVIEDTGGSNLWVPSVYCSSPACSNRNRENPOOSYTGCTNQKLSV 133
QY 133 PYTGKWEGLGTLVSLPHGPNVTVRANITAAITESDK-FFINGSNWEGILGLAYAEIAR 191
DB 134 AVGTGSMTGILGYDTVQV---CGITDITNQIFGLSETEPGSFYIYAPDGLGLAYPSIA- 189
QY 192 PDSLEPFEDSLVKQTHV-PNLFSLQLCGAGFPLNQSEVLASVSGSMII--GGIDHLSLYTG 249
DB 190 -SSGATPVEDNIWNLQVLSQDLFSVYLS-----NDQGGVWFMGGIDSSVFTG 237
QY 250 SLWYTPIRREWEYVEIIVRVEINGQDLKM--DCKEYNDKSIIVDSGTTNLRPKKVFEEA 307
DB 238 NLNMVPLSSSTYQWITVDSITMNGQVIACSGC-----QAIVDTGTSLLSGPTNAI-AS 290
QY 308 VKSIIKAASSTKFPDGFWLGEQLV-CWQAGTTPWNIPFVVISLVMGEVNTNQSFRTILPQ 366
DB 291 IOGYIGASQANAN-----GEMVYVSCSAINTLNIPV-----TINGV 325
QY 367 QY-LRPVEDVATSDCC---YKFAISOSTGT--VMGAVIMEGFGYVVFDRARKRIGFA 418
DB 326 QYPLPFSAYVLSQSQGCTSGFGQMDIPTSSGELWILGDVFIQYFTVFDGRNNQVGLA 383

RESULT 15
Q9GMV8

ID Q9GMV8 PRELIMINARY; PRT; 387 AA.
AC Q9GMV8;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE PEPINOGEN A.
GN PGNA.
OS Sorex unguiculatus (Long-clawed shrew).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Insectivora; Soricidae; Sorex.
OX NCBI_TaxID=62275;
RN [1]
RA Narita Y., Oda S., Takenaka O., Kageyama T.;
RT "Phylogenetic position of Insectivora inferred from the cDNA sequences
of pepsinogen A and C.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB047244; BAB11750.1; -.
SQ SEQUENCE 387 AA; 41514 MW; F2EB2E331FAA248F CRC64;

```

Query Match      11.5%; Score 305; DB 6; Length 387;
Best Local Similarity 28.0%; Pred. No. 1.9e-16;
Matches 100; Conservative 64; Mismatches 131; Indels 62; Gaps 15;

QY 77 YVEMTVGSPQTLNILDVGTSSNFAVG-----AAPHPELHRYQRLSSTYRDLRKGVYV 132
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 75 YFGTISIGTPPOEFTVIFDTGSSNLWVPSIYCSPACSNHNRFPDQKSTFKPTQTVSI 134
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 133 PYTGKWEGLCTDLVSPHGPVNVIVRANIAAITESDK-FFINGSNWEGILGLAYAEIAR 191
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 135 AYTGSMTGVLGYDRVQV---AGIADTNQIFGLSQSEPGFLYSPFDGILGLAYPSIS- 190
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 192 PDSLEPRFDSLKVQTHV-PNLFSLQLCGAGPPLNQSEVLASVGGSMIIGGDHSLYTGS 250
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 191 -SSGATPVFDNNWNOGLVSDQLFSVILSN-----DQS-----GSVMFEGGIDSSYYTGS 239
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 251 LMWTFIRREWEYEVILIVRVEINGODLKMD--CKEYNYDKSIYDSGTTNLRPKKVFEEAV 308
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 240 LNWVPLSSEGYWQITVDSITMNGQSIACNGGC-----QAIVDGTGTLSSGPTNAIANIQ 293
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 309 KSIKAASSTKFPDGFWLGEOLV-CWQAGTTPWNIFPVLSLYLMGEVINOSEFRITILPQQ 367
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 294 SKIGASQNSQ-----COMAVSCSSTKNLPDIVF-----TINGIQ 327
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 368 YLRPVED-VATSQDDC---YKFAISQSSTGT--VMGAVIMEGFYVVFDRARKRIGFA 418
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 328 YPLPASAYILQSQEGCGSGFGQMDIPTSSGELWILGDVFIRQYFTVFDNRANNOVGLA 384

```

Search completed: September 6, 2001, 16:49:45
Job time: 729 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:43:26 ; Search time 225.25 Seconds
(without alignments)
2.691 Million cell updates/sec

Title: US-09-603-713-4

Perfect score: 49

Sequence: 1 SEVKMDAERF 10

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_0601.*

1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49	100.0	10	13 AAR24261	Human amyloidin pr
2	49	100.0	10	21 AAY69703	Beta-APP alpha-sec
3	49	100.0	10	22 AAB6574	Synthetic peptide
4	49	100.0	10	22 AAB65208	Human APP derived
5	49	100.0	10	22 AAB61336	Synthetic peptide f
6	49	100.0	13	19 AAW70869	Beta-amyloid pepti
7	49	100.0	16	21 AAB06315	Human beta-amyloid
8	49	100.0	16	21 AAB06317	Human beta-amyloid
9	49	100.0	20	21 AAY69713	Beta-APP alpha-sec
10	49	100.0	33	20 AAW98002	Amyloid precursor
11	49	100.0	39	21 AAY69717	Beta-APP alpha-sec

12	49	100.0	45	18 AAW26512	Amyloid precursor
13	49	100.0	45	18 AAW26392	Amyloid precursor
14	49	100.0	45	19 AAW44748	APP-REP 751 [BAP d
15	49	100.0	45	19 AAW42977	Deletion beta-amy1
16	49	100.0	53	16 AAR64168	Variant beta amylo
17	49	100.0	54	21 AAB32126	Amyloid-beta precu
18	49	100.0	57	21 AAB10910	Human amyloid prec
19	49	100.0	58	15 AAR58937	Amyloid precursor
20	49	100.0	63	18 AAW26511	Amyloid precursor
21	49	100.0	63	18 AAW26391	Amyloid precursor
22	49	100.0	63	19 AAW44747	APP-REP 751 [BAP p
23	49	100.0	63	19 AAW44746	APP-REP 751 [BAP E
24	49	100.0	63	19 AAW42975	Beta-amyloid pepti
25	49	100.0	63	19 AAW42976	Beta-amyloid pepti
26	49	100.0	67	19 AAW71377	Peptide derived fr
27	49	100.0	104	19 AAW51100	Amino acid sequenc
28	49	100.0	112	17 AAR93556	Familial Alzheimer
29	49	100.0	115	20 AAW97999	London-PAD APP pol
30	49	100.0	117	19 AAW51102	Flag-amyloid prote
31	49	100.0	162	9 AAR83151	Deduced sequence i
32	49	100.0	162	12 AAR10023	Beta-amyloid-relat
33	49	100.0	162	14 AAR37863	Deduced from clone
34	49	100.0	249	15 AAR65798	Beta-amyloid precu
35	49	100.0	264	10 AAP90497	Protein sequence i
36	49	100.0	264	10 AAP90609	Sequence of amy 37
37	49	100.0	487	18 AAW26510	Amyloid precursor
38	49	100.0	487	18 AAW26394	Amyloid precursor
39	49	100.0	487	19 AAW44745	APP-REP 751 protei
40	49	100.0	487	19 AAW42979	Amyloid precursor
41	49	100.0	492	14 AAR45229	APP-REP 751 amylo1
42	49	100.0	492	18 AAW26509	Amyloid precursor
43	49	100.0	492	18 AAW26393	Amyloid precursor
44	49	100.0	492	19 AAW44744	APP-REP 751 protei
45	49	100.0	492	19 AAW42978	Amyloid precursor

ALIGNMENTS

RESULT 1
AAR24261
ID AAR24261 standard; Protein; 10 AA.
XX
AC AAR24261;
XX 09-NOV-1992 (first entry)
DT Human amyloidin protease substrate sequence #1.
DE Alzheimer's disease; beta amyloid precursor protein; APP; zinc;
KW metalloprotease; hAP; protease inhibitor; APP592-601
KW Synthetic.
OS
XX Key Location/Qualifiers
FH Modified-site 1 /note- "Acetylated-Ser"
FT
XX WO9207068-A.
XX
XX 30-APR-1992.
XX
XX 04-OCT-1991; 91WO-US07290.
XX
XX 05-OCT-1990; 90US-0594122.
XX 30-SEP-1991; 91US-0766351.
XX
XX (ATHE-) ATHENA NEUROSCIENCES INC.
XX (ELIL) LILLY & CO ELI.
XX Dovey HF, Johnstone EM, Little SP, McConlogue L, Seubert PA;
XX Sinha S;

DR WPI; 1992-167148/20.
 XX Human amyloidin protease - used for cleaving Met-Asp bond in
 PT amyloid-like substrate for identifying protease inhibitors
 XX
 XX
 XX Claim 1; Page 52; 62pp; English.
 XX
 CC Claimed human amyloidin protease is defined by its ability to
 CC cleave the Met-Asp bond of this synthetic substrate. The substrate,
 CC which corresponds to residues 592 to 601 of the 695 amino acid APP,
 CC can be used in an assay for identifying inhibitors of proteases
 CC which cleave Met-Asp bonds, e.g. amyloidin, human skin chymase or
 CC rat mast cell protease I or II.
 CC See AAR24260-3, AAR24266-7 and AAQ24875-Q24887.
 XX
 XX Sequence 10 AA;
 SQ

Query Match 100.0%; Score 49; DB 13; Length 10;
 Best Local Similarity 100.0%; Pred. No. 0.00022;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SEVKMDAEFR 10
 DB 1 sevkmdaeifr 10
 |||||

RESULT 2
 AAY69703
 ID AAY69703 standard; peptide; 10 AA.
 AC
 XX
 XX
 XX 11-APR-2000 (first entry)
 DT
 DE Beta-APP alpha-secretase substrate [KMD]-APP(-5,+5).
 XX Nootropic; neuroprotective; beta-amyloid precursor protein; metabolism;
 KW cleavage site; beta-secretase; neurodegenerative disease;
 KW Alzheimer's disease.
 XX Homo sapiens.
 OS
 XX WO9964587-A1.
 PN
 XX 16-DEC-1999.
 PD
 XX 04-JUN-1999; 99WO-FR01326.
 PF
 XX 05-JUN-1998; 98FR-0007068.
 PR
 XX 31-MAR-1999; 99US-0122599.
 PR
 XX (RHON) RHONE-POULENC RORER SA.
 PA (UYPA-) UNIV CURIE PARIS VI P & M.
 XX
 XX Rhoham M, Munoz-Gimenez N, Moutaouakil M, Cohen P, Bertrand P;
 PI
 XX WPI; 2000-097537/08.
 DR
 XX Polypeptide with beta-secretase activity, specific for wild-type
 PT amyloid precursor protein, useful in treating Alzheimer's disease
 PT
 XX Example 3; Page 24; 44pp; French.
 PS
 XX Peptides AAY69702-Y69718 represent synthetic peptide substrates for a
 CC novel polypeptide with beta-secretase activity that can cleave
 CC specifically the natural beta-amyloid precursor protein (bAPP). Normal
 CC cleavage of the protein occurs between amino acids Met596-Asp597 and
 CC Val636-Ile637 (positions 4-5 and 44-45 of AAY69701). The novel
 CC polypeptide is used to identify agents that interact specifically with
 CC it. These agents regulate metabolism of APP, particularly they slow down
 CC or reduce production of beta-amyloid, so can be used to treat
 CC neurodegenerative diseases, particularly Alzheimer's disease.
 PC

Query Match 100.0%; Score 49; DB 21; Length 10;
 Best Local Similarity 100.0%; Pred. No. 0.00022;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SEVKMDAEFR 10
 DB 1 sevkmdaeifr 10
 |||||

RESULT 3
 AAB66574
 ID AAB66574 standard; Peptide; 10 AA.
 AC
 XX
 XX
 XX 12-APR-2001 (first entry)
 DT
 DE Synthetic peptide derived from APP beta-secretase site.
 XX
 XX Memapsin 2; nootropic; neuroprotective; amyloid
 KW APP; memapsin 2 inhibitor; Alzheimer's disease.
 XX
 OS Synthetic.
 XX
 XX WO200100665-A2.
 PN
 XX 04-JAN-2001.
 PD
 XX 27-JUN-2000; 2000WO-US17742.
 PF
 XX 28-JUN-1999; 99US-0141363.
 PR
 XX 30-NOV-1999; 99US-0168060.
 PR
 XX 25-JAN-2000; 2000US-0177836.
 PR
 XX 27-JAN-2000; 2000US-0178368.
 PR
 XX 08-JUN-2000; 2000US-0210292.
 PR
 XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.
 PA (UNII) UNIV ILLINOIS FOUND.
 PA
 XX Tang JJN, Hong L, Ghosh AK;
 PI
 XX WPI; 2001-137933/14.
 DR
 XX Novel memapsin 2 inhibitors which bind to active site of memapsin 2
 PT having 2 catalytic aspartic residues and substrate binding cleft, used
 PT to treat Alzheimer's disease by blocking amyloid precursor protein
 PT cleavage
 PT
 XX Disclosure; Page 11; 86pp; English.
 PS
 XX The present sequence is given in a specification relating to an inhibitor
 CC of catalytically active memapsin 2. The inhibitor binds to the memapsin 2
 CC active site, which is defined by the presence of two catalytic aspartic
 CC residues and a substrate binding cleft. The inhibitor is useful for
 CC the treatment and diagnosis of Alzheimer's disease. It is useful in
 CC screens for individuals with a genetic predisposition to Alzheimer's
 CC disease. The inhibitor is useful as a reagent for specifically binding to
 CC memapsin 2 or memapsin 2 analogues and for aiding in memapsin 2
 CC isolation, purification and characterisation.
 XX
 XX Sequence 10 AA;
 SQ

Query Match 100.0%; Score 49; DB 22; Length 10;
 Best Local Similarity 100.0%; Pred. No. 0.00022;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SEVKMDAEFR 10
 DB 1 sevkmdaeifr 10
 |||||

Db 1 sevkmdae fr 10

RESULT 4

AAB46208
ID AAB46208 standard; peptide; 10 AA.

XX AC AAB46208;

XX DT 04-APR-2001 (first entry)

XX DE Human APP derived immunogenic peptide #4.

XX KW Amyloid deposit; APP; Abeta; brain; human; clearing response; nootropic;
XX KW Fc receptor mediated phagocytosis; immunogenic response; neuroprotective;
XX KW amyloid precursor protein; Alzheimer's disease.

XX OS Homo sapiens.

XX PN WO200072880-A2.

XX PD 07-DEC-2000.

XX PF 26-MAY-2000; 2000WO-US14810.

XX PR 28-MAY-1999; 99US-0322289.

XX PA (NEUR-) NEURALAB LTD.

XX PI Schenk DB, Bard F, Vasquez NJ, Yednock T;

XX DR WPI; 2001-032104/04.

XX PT Preventing or treating a disease associated with amyloid deposits,
XX PT especially Alzheimer's disease, comprises administering amyloid
XX PT specific antibody

XX PS Disclosure; Figure 19; 143pp; English.

XX CC This invention describes a novel method of preventing or treating a
XX CC disease associated with amyloid deposits of amyloid precursor protein
XX CC (APP) Abeta fragments in the brain of a patient, which comprises
XX CC administering to the patient: (a) an antibody that binds to Abeta, the
XX CC antibody binds to an amyloid deposit and induces a clearing response (Fc
XX CC receptor mediated phagocytosis) against it (b) a polypeptide containing
XX CC an N-terminal segment of at least residues 1-5 of Abeta; or (c) an agent
XX CC that induces an immunogenic response against residues 1-3 to 7-11 of
XX CC Abeta. The products of the invention have nootropic and neuroprotective
XX CC activity. The method is also useful for monitoring a course of treatment
XX CC being administered to a patient e.g. active and passive immunization. The
XX CC methods are useful for prophylactic and therapeutic treatment of
XX CC Alzheimer's disease.

XX SQ Sequence 10 AA;

Query Match 100.0%; Score 49; DB 22; Length 10;

Best Local Similarity 100.0%; Pred. No. 0.00022;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SEVKMDAEFR 10

Db 1 sevkmdae fr 10

RESULT 5

AAB61336
ID AAB61336 standard; peptide; 10 AA.

XX AC AAB61336;

XX DT 02-APR-2001 (first entry)

XX

DE Sythetic peptide from beta amyloid precursor protein.

XX Memapsin 2; catalyst; Alzheimer's.

XX OS Unidentified.

XX PN WO200100663-A2.

XX PD 04-JAN-2001.

XX PF 27-JUN-2000; 2000WO-US17661.

XX PR 28-JUN-1999; 99US-0141363.

XX PR 30-NOV-1999; 99US-0168060.

XX PR 25-JAN-2000; 2000US-0177836.

XX PR 27-JAN-2000; 2000US-0178368.

XX PR 08-JUN-2000; 2000US-0210292.

XX PA (OKLA-) OKLAHOMA MEDICAL RES FOUND.

XX PI Tang JJN, Lin X, Koelsch G;

XX DR WPI; 2001-102885/11.

XX PT Purified recombinant catalytically active memapsin 2, used to screen
XX PT inhibitors of it, which are used to treat and prevent Alzheimer's
XX PT disease

XX PS Claim 6; Page 11; 86pp; English.

XX CC The present invention relates to a purified recombinant
XX CC catalytically active memapsin 2. The invention may be used for
XX CC isolating inhibitors which are used to treat or prevent
XX CC Alzheimer's disease. The invention may also be used to screen
XX CC for individuals more genetically prone to develop Alzheimer's
XX CC disease.

XX SQ Sequence 10 AA;

Query Match 100.0%; Score 49; DB 22; Length 10;

Best Local Similarity 100.0%; Pred. No. 0.00022;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SEVKMDAEFR 10

Db 1 sevkmdae fr 10

RESULT 6

AAW70869

ID AAW70869 standard; peptide; 13 AA.

XX AC AAW70869;

XX DT 04-FEB-1999 (first entry)

XX DE Beta-amyloid peptide to create a monoclonal antibody.

XX KW Beta-amyloid precursor protein; beta-APP; beta-amyloid peptide;
XX KW antibody; amyloid deposit; Alzheimer's disease.

XX OS Synthetic.

XX OS Homo sapiens.

XX PN WO9844955-A1.

XX PD 15-OCT-1998.

XX PF 09-APR-1998; 98WO-US06900.

XX PR 09-APR-1997; 97US-0041850.

XX

PA (MCIN/) MCINNIS P.A.
 PA (MIND-) MINDSET LTD.
 XX Chain DG;
 XX WPI; 1998-594476/50.
 DR Preventing or inhibiting progression of Alzheimer's Disease -
 XX comprises use of recombinant DNA encoding an antibody specific for
 PT the N- or C-terminus of an amyloid-beta peptide
 XX
 PS Example 1; Page 47; 58pp; English.
 XX
 CC The present sequence represents a peptide derived from beta-amyloid
 CC precursor protein (beta-APP). The peptide is a beta-amyloid
 CC peptide and is used to produce a monoclonal antibody. The specification
 CC describes a method for prevention or inhibition of progression of
 CC Alzheimer's disease. The method comprises administering a composition
 CC comprising a recombinant DNA molecule containing a gene encoding a
 CC recombinant antibody end-specific for the N-terminus or the C-terminus
 CC of an amyloid-beta peptide, operably linked to a promoter which is
 CC expressed in the central nervous system. The recombinant antibody
 CC molecules prevent the accumulation of beta-amyloid peptides in the
 CC extracellular space, interstitial fluid and cerebrospinal fluid and the
 CC aggregation of such peptides into amyloid deposits in the brain. They
 CC also inhibit the progression of Alzheimer's disease by inhibiting the
 CC interaction of beta-amyloid peptides mediating Alzheimer's disease
 CC induced neurotoxicity and inhibiting the Alzheimer's disease induced
 CC complement activation and cytokine release involved in the inflammatory
 CC process.
 XX
 SQ Sequence 13 AA;

Query Match 100.0%; Score 49; DB 19; Length 13;
 Best Local Similarity 100.0%; Pred. No. 0.00029;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SEVKMDAEFR 10
 |||||
 Db 3 sevkmdaefr 12

RESULT 7
 AAB06315
 ID AAB06315 standard; peptide; 16 AA.
 AC AAB06315;
 XX
 XX
 DT 03-OCT-2000 (first entry)
 XX
 DE Human beta-amyloid precursor protein beta-secretase cleavage site.
 XX
 KW Human; beta-amyloid precursor protein; beta-APP; beta-secretase;
 KW subtilisin-kexin isoenzyme 1; SKI-1;
 KW pro-brain-derived neurotrophic factor; proBDNF; antilipaeamic;
 KW cytostatic; vasotropic; SKI-1 inhibitor; hypercholesterolaemia;
 KW liver steatosis; Ras-dependent cancer; restenosis;
 KW amyloid protein formation.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH Cleavage-site 8..9
 FT WO200026348-A2.
 XX
 PN 11-MAY-2000.
 XX
 PD 04-NOV-1999; 99WO-CA01058.
 XX
 PF 04-NOV-1998; 98CA-2249648.
 XX
 PR 04-NOV-1998; 98CA-2249648.
 XX
 PX

(RECL-) INST RECH CLINIQUES MONTREAL.

Seidah N, Chretien M, Marcinkiewicz M, Laaksonen R, Davignon J;
 WPI; 2000-365601/31.

Novel soluble proteic fragment of subtilisin-kexin isoenzyme for
 PT producing a polypeptide useful for treating hypercholesterolemia, liver
 PT steatosis and amyloidosis, comprises a specific amino acid sequence -
 XX
 PS Example 4; Page 51; 119pp; English.

XX The present sequence is the beta-secretase site of human beta-amyloid
 CC precursor protein (beta-APP). The sequence may be cleaved
 CC by a mammalian subtilase, specifically subtilisin-kexin isoenzyme 1
 CC (SKI-1), a type-1 membrane-bound proteinase. Peptides which bind to and
 CC are cleaved by SKI-1 may be used for monitoring SKI-1 activity, for
 CC screening inhibitors of SKI-1 activity, or for screening enhancers of
 CC SKI-1 activity. Proteic fragments of SKI-1 which bind to the SKI-1
 CC catalytic site may be used as inhibitors of SKI-1 activity. They may be
 CC used to treat diseases involving overexpression of SKI-1 or SKI-1
 CC substrate. Such diseases include hypercholesterolaemia, high levels of
 CC fatty acids, lipids or farnesyl pyrophosphate, liver steatosis,
 CC Ras-dependent cancer, restenosis and amyloid protein formation.

Sequence 16 AA;

Query Match 100.0%; Score 49; DB 21; Length 16;
 Best Local Similarity 100.0%; Pred. No. 0.00037;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SEVKMDAEFR 10
 |||||
 Db 4 sevkmdaefr 13

RESULT 8
 AAB06317
 ID AAB06317 standard; peptide; 16 AA.
 XX
 XX AAB06317;
 AC

DT 03-OCT-2000 (first entry)

Human beta-amyloid precursor protein beta-epsilon1-secretase site.

Human; beta-amyloid precursor protein; beta-APP;
 KW beta-epsilon1-secretase; subtilisin-kexin isoenzyme 1; SKI-1;
 KW pro-brain-derived neurotrophic factor; proBDNF; antilipaeamic;
 KW cytostatic; vasotropic; SKI-1 inhibitor; hypercholesterolaemia;
 KW liver steatosis; Ras-dependent cancer; restenosis;
 KW amyloid protein formation.

XX Homo sapiens.

XX Key Location/Qualifiers
 FH Cleavage-site 8..9

XX WO200026348-A2.

XX 11-MAY-2000.

XX 04-NOV-1999; 99WO-CA01058.

XX 04-NOV-1998; 98CA-2249648.

(RECL-) INST RECH CLINIQUES MONTREAL.

Seidah N, Chretien M, Marcinkiewicz M, Laaksonen R, Davignon J;
 WPI; 2000-365601/31.

PT Novel soluble proteic fragment of subtilisin-kexin isoenzyme for
 PT producing a polypeptide useful for treating hypercholesterolemia, liver
 PT steatosis and amyloidosis, comprises a specific amino acid sequence -
 XX
 PS Example 4; Page 51; 119pp; English.

XX The present sequence is the beta-epsilon1-secretase site of human
 CC beta-amyloid precursor protein (beta-APP). The sequence may be cleaved
 CC by a mammalian subtilase, specifically subtilisin-kexin isoenzyme 1
 CC (SKI-1), a type-1 membrane-bound proteinase. Peptides which bind to and
 CC are cleaved by SKI-1 may be used for monitoring SKI-1 activity, for
 CC screening inhibitors of SKI-1 activity, or for screening enhancers of
 CC SKI-1 activity. Peptide fragments of SKI-1 which bind to the SKI-1
 CC catalytic site may be used as inhibitors of SKI-1 activity. They may be
 CC used to treat diseases involving overexpression of SKI-1 or SKI-1
 CC substrate. Such diseases include hypercholesterolemia, high levels of
 CC fatty acids, lipids or farnesyl pyrophosphate, liver steatosis,
 CC Ras-dependent cancer, restenosis and amyloid protein formation.

XX Sequence 16 AA;

Query Match 100.0%; Score 49; DB 21; Length 16;
 Best Local Similarity 100.0%; Pred. No. 0.00037; Indels 0; Gaps 0;
 Matches 10; Conservative 0; Mismatches 0;

Qy 1 SEVKMDAEFR 10
 Db 2 sevkmdaeifr 11
 |||||

RESULT 9

AAV69713
 ID AAV69713 standard; peptide: 20 AA.

XX AC AAV69713;

XX DT 11-APR-2000 (first entry)

XX DE Beta-APP alpha-secretase substrate [KM]-APP(-10,+10).

XX KW Nootropic; neuroprotective; beta-amyloid precursor protein; metabolism;
 KW cleavage site; beta-secretase; neurodegenerative disease;
 KW Alzheimer's disease.

XX OS Homo sapiens.

XX PN W09964587-A1.

XX PD 16-DEC-1999.

XX PF 04-JUN-1999; 99WO-FR01326.

XX PR 05-JUN-1998; 98FR-0007068.

XX PR 31-MAR-1999; 99US-0122599.

XX PA (RHON) RHONE-POULENC RORER SA.

XX PA (UYPA-) UNIV CURIE PARIS VI P & M.

XX PI Rholam M, Munoz-Gimenez N, Moutaouakil M, Cohen P, Bertrand P;

XX DR WPI; 2000-097537/08.

XX PT Polypeptide with beta-secretase activity, specific for wild-type

XX PT amyloid precursor protein, useful in treating Alzheimer's disease -

XX PS Example 3; Page 24; 44pp; French.

XX CC Peptides AAV69702-Y69718 represent synthetic peptide substrates for a
 CC novel polypeptide with beta-secretase activity that can cleave
 CC specifically the natural beta-amyloid precursor protein (bAPP). Normal
 CC cleavage of the protein occurs between amino acids Met596-Asp597 and
 CC Val636-Ile637 (positions 4-5 and 44-45 of AAV69701). The novel

CC polypeptide is used to identify agents that interact specifically with
 CC it. These agents regulate metabolism of APP, particularly they slow down
 CC or reduce production of beta-amyloid, so can be used to treat
 CC neurodegenerative diseases, particularly Alzheimer's disease.

SQ Sequence 20 AA;

Query Match 100.0%; Score 49; DB 21; Length 20;
 Best Local Similarity 100.0%; Pred. No. 0.00048;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SEVKMDAEFR 10
 Db 6 sevkmdaeifr 15
 |||||

RESULT 10

AAW98002
 ID AAW98002 standard; Protein; 33 AA.

XX AC AAW98002;

XX DT 21-JUN-1999 (first entry)

XX DE Amyloid precursor protein (aa656-678) with Swedish mutation.

XX KW Amyloid precursor protein; APP; human; gene targeting;
 KW homologous recombination; transgenic mouse; transgenic animal;
 KW animal model; Alzheimer's disease.

XX OS Mus musculus.

XX PN W09909150-A1.

XX PD 25-FEB-1999.

XX PF 18-AUG-1997; 97WO-US14507.

XX PR 18-AUG-1997; 97WO-US14507.

XX PA (FARB) BAYER CORP.

XX PI Wirak DO;

XX DR WPI; 1999-181029/15.

XX PT Modification of target nucleic acids - by homologous recombination,
 PT used particularly for introducing a humanised amyloid precursor
 PT protein gene into rodents for producing models of Alzheimer's
 PT disease

XX PS Disclosure; Page 145; 209pp; English.

XX CC This polypeptide comprises residues 656-678 of a murine amyloid
 CC precursor protein (APP). The invention provides a novel gene
 CC targeting strategy that facilitates the introduction of one or
 CC more specific mutations into any gene in a single double reciprocal
 CC homologous recombination step. The method has been used
 CC particularly for introducing a humanised APP gene into rodents for
 CC producing animal models of Alzheimer's disease (AD). 4 Independent
 CC lines of transgenic mice (lines ES5007, ES5103, ES5401 and ES5403)
 CC have been created using the gene targeting technique applied to
 CC embryonic stem cells. In each line, the mouse APP gene was modified
 CC to encode a mouse/human hybrid (m/hAPP) where amino acid residues
 CC 666-770 of APP770 were encoded by human cDNA sequences instead of
 CC mouse genomic exons (exons 16-18). Within these residues, only 3
 CC amino acid differences exist between the mouse and human proteins,
 CC i.e. Gly-676 to Arg, Phe-681 to Thr and Arg-684 to His. The
 CC exon-cDNA fusion gene therefore encodes an APP containing a
 CC humanised beta-amyloid domain. Swedish- and/or London-FAD APP
 CC mutations have also been introduced (see also AAW97997-W98001).

XX

2


```

XX AC AAW26392;
XX DT 15-DEC-1997 (first entry)
XX DE Amyloid precursor protein fragment APP-REP 751 (BAP del11-28).
XX KW Amyloid precursor protein; APP; beta-amyloid protein; BAP;
XX KW substrate; muten; secretase; Alzheimer's disease; human.
XX OS Chimeric Homo sapiens;
XX OS Chimeric synthetic.
XX FH Key Location/Qualifiers
XX FT Cleavage-site 7..8 /note= "secretase cleavage site"
XX FT Peptide 10..33 /label= BAP(del11-28)
XX FT Domain 20..42 /note= "truncated beta-amyloid protein"
XX FT /label= Transmembrane
XX PN US5652092-A.
XX PD 29-JUL-1997.
XX PF 01-MAY-1992; 92US-0877675.
XX PR 20-SEP-1993; 93US-0123659.
XX PR 01-MAY-1992; 92US-0877675.
XX PR 05-JUN-1995; 95US-0462859.
XX PA (AMCY ) AMERICAN CYANAMID CO.
XX PI Jacobsen JS, Vitek MP;
XX PI WPI: 1997-392937/36.
XX DR Screening for compounds which reduce beta-amyloid protein formation
XX PT - using cells which express a construct encoding a marker and an
XX PT amyloid precursor muten derived from APP isoforms
XX PS Disclosure: Fig 5A; 84pp; English.
XX CC This peptide sequence shows the region of amyloid precursor protein
XX CC (APP) that includes a truncated beta-amyloid protein (BAP) lacking
XX CC the native secretase cleavage/recognition site. In an attempt to
XX CC engineer an APP non-cleavable substrate for secretase, an
XX CC APP-reporter (APP-REP) protein that carries the BAP deletion has
XX CC been expressed in recombinant host cells. Deletion of these 18
XX CC amino acids, however, still resulted in the secretion of an
XX CC N-terminal APP-reporter fragment into the cytoplasm. Non-
XX CC cleavable APP substrates can be used to detect other putative
XX CC abnormal APP processing events. They can also be used to
XX CC investigate cellular post-translational modifications to APP in
XX CC order to determine the potential influence on normal secretase and
XX CC abnormal BAP 'clipping' activities.
XX SQ Sequence 45 AA;

Query Match 100.0%; Score 49; DB 18; Length 45;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SEVKMDAEFR 10
Db 5 sevkmdaeFr 14

RESULT 14
AAW44748
ID AAW44748 standard; Protein; 45 AA.

```

```

XX AC AAW44748;
XX DT 01-JUN-1998 (first entry)
XX DE APP-REP 751 [BAP delta(11-28)] peptide.
XX KW Amyloid precursor protein; APP; APP 751 isoform; deletion; substrate P;
XX KW epitope; Met-enkephalin; detection; secretase; beta-amyloid protein; BAP;
XX KW Alzheimer's disease; cleavage.
XX OS Homo sapiens.
XX OS Synthetic.
XX FH Key Location/Qualifiers
XX FT Cleavage-site 7..8 /note= "putative secretase cleavage site"
XX FT Misc-difference 19..20 /note= "residues 11-28 of the wild type BAP sequence
XX FT are deleted from between these positions"
XX PN US5693478-A.
XX PD 02-DEC-1997.
XX PF 05-JUN-1995; 95US-0464247.
XX PR 20-SEP-1993; 93US-0123659.
XX PR 01-MAY-1992; 92US-0877675.
XX PR 05-JUN-1995; 95US-0464247.
XX PA (AMCY ) AMERICAN CYANAMID CO.
XX PI Jacobsen JS, Vitek MP;
XX PI WPI: 1998-031744/03.
XX DR Amyloid precursor muten reporter molecule assay containing antibody
XX PT recognised marker - used to study pathways associated with
XX PT Alzheimer's disease
XX PS Disclosure: Fig 5A; 84pp; English.
XX CC This sequence represent the beta-amyloid protein sequence from the
XX CC construct APP-REP751 [BAP delta(11-28)]. The mutant sequence contains
XX CC a deletion of the wild type BAP residues 11-28. This causes a
XX CC shortening of the BAP sequence. This may affect cleavage of the BAP by
XX CC the "secretase" dependent on whether the "secretase" recognises the
XX CC cleavage site by a positional effect or by sequence. The mutant sequence
XX CC can be used in a method to study secretase and beta-amyloid protein
XX CC (BAP)-generating pathways associated with Alzheimer's disease by
XX CC studying proteolytic cleavage of the reporter polypeptides (e.g.
XX CC AAW44744 and AAW44745).
XX SQ Sequence 45 AA;

Query Match 100.0%; Score 49; DB 19; Length 45;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SEVKMDAEFR 10
Db 5 sevkmdaeFr 14

RESULT 15
AAW42977
ID AAW42977 standard; peptide; 45 AA.
XX AC AAW42977;
XX DT 01-MAY-1998 (first entry)

```

XX DE Deletion beta-amyloid peptide (BAP) derived from APP-REP 751.
XX KW Beta-amyloid peptide; BAP; extracellular BAP plaque;
KW cerebrovascular deposit; Alzheimers disease; Downs syndrome;
KW amyloid precursor protein; APP; secretase; BAP aggregation;
KW abnormal proteolytic cleavage.
XX OS Homo sapiens.
XX PH Key Location/Qualifiers
FT Domain 20..43
FT /note- "putative transmembrane domain"
XX US5703209-A.
XX PD 30-DEC-1997.
XX PF 05-JUN-1995; 95US-0464248.
XX PR 20-SEP-1993; 93US-0123659.
XX PR 01-MAY-1992; 92US-0877675.
XX PA (AMCY) AMERICAN CYANAMID CO.
XX PI Jacobsen JS, Vitek MP;
XX WPI; 1998-076482/07.
XX PT Amyloid precursor protein fusion polypeptides - comprising APP
PT fragment and marker, useful for research and drug screening
XX PS Disclosure: Fig 5A; 84pp; English.
XX CC The present sequence represents a beta-amyloid peptide (BAP), with
CC a deletion amino acids 11-28 (numbered according to AAW42976). Abnormal
CC accumulation of extracellular BAP in plaques and cerebrovascular
CC deposits is characteristic in brains of individuals suffering from
CC Alzheimers disease and Downs syndrome. BAP is a poorly soluble,
CC self-aggregating protein which is derived from a larger amyloid precursor
CC protein (APP). APP is expressed as an integral membrane protein, and is
CC cleaved by secretase, between BAP 16lys and 17Leu. Cleavage at this site
CC precludes amyloidogenesis and results in the release of the
CC amino-terminal APP fragment. Three major isoforms of APP exist: APP-695,
CC APP-751 and APP-770. These isoforms are derived by alternative splicing.
CC APP-ARP 751 is a deletion construct of APP-751, which has a deletion of
CC 276 amino acids to within 15 amino acids of the BAP domain. APP can be
CC used as a substrate for studying abnormal proteolytic cleavage which
CC results in the release of BAP, and also to screen for drugs that will
CC inhibit such cleavage.
XX SQ Sequence 45 AA;

Query Match 100.0%; Score 49; DB 19; Length 45;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVKMDAEFR 10
Db |
5 sevkmdaefr 14

Search completed: September 6, 2001, 16:43:27
Job time: 356 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run On: September 6, 2001, 16:39:34 ; Search time 113.12 Seconds
(without alignments)
1.820 Million cell updates/sec

Title: US-09-603-713-4

Perfect score: 49

Sequence: 1 SEVKMDAEFR 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49	100.0	11	5	PCT-US94-07043A-7
2	49	100.0	12	5	PCT-US94-07043A-2
3	49	100.0	16	5	PCT-US94-07043A-1
4	49	100.0	27	1	US-08-141-324-11
5	49	100.0	27	1	US-08-541-902-11
6	49	100.0	45	1	US-08-462-859A-5
7	49	100.0	45	1	US-08-123-659A-5
8	49	100.0	45	1	US-08-464-247A-5
9	49	100.0	45	1	US-08-464-248A-5
10	49	100.0	58	1	US-08-371-930-25
11	49	100.0	58	5	PCT-US94-01712-25
12	49	100.0	63	1	US-08-462-859A-3
13	49	100.0	63	1	US-08-462-859A-4
14	49	100.0	63	1	US-08-123-659A-3
15	49	100.0	63	1	US-08-123-659A-4
16	49	100.0	63	1	US-08-464-247A-3
17	49	100.0	63	1	US-08-464-247A-4
18	49	100.0	63	1	US-08-464-248A-3
19	49	100.0	63	1	US-08-464-248A-4
20	49	100.0	105	2	US-08-729-345-1
21	49	100.0	117	2	US-08-729-345-3
22	49	100.0	152	6	5187153-4
23	49	100.0	162	6	5220013-4
24	49	100.0	162	6	5223482-4
25	49	100.0	264	1	US-07-990-893-5
26	49	100.0	487	1	US-08-462-859A-9
27	49	100.0	487	1	US-08-123-659A-9

Query Match

100.0% ; Score 49; DB 5; Length 11;

28	49	100.0	487	1	US-08-464-247A-9	Sequence 9, Appl
29	49	100.0	487	1	US-08-464-248A-9	Sequence 9, Appl
30	49	100.0	492	1	US-08-462-859A-7	Sequence 7, Appl
31	49	100.0	492	1	US-08-123-659A-7	Sequence 7, Appl
32	49	100.0	492	1	US-08-464-247A-7	Sequence 7, Appl
33	49	100.0	492	1	US-08-464-248A-7	Sequence 7, Appl
34	49	100.0	537	1	US-08-453-552-4	Sequence 4, Appl
35	49	100.0	537	1	US-08-710-637-4	Sequence 4, Appl
36	49	100.0	537	5	PCT-US93-00907-4	Sequence 4, Appl
37	49	100.0	656	1	US-08-371-930-23	Sequence 23, Appl
38	49	100.0	656	5	PCT-US94-01712-23	Sequence 23, Appl
39	49	100.0	676	1	US-08-371-930-24	Sequence 24, Appl
40	49	100.0	676	5	PCT-US94-01712-24	Sequence 24, Appl
41	49	100.0	694	1	US-08-339-152A-18	Sequence 18, Appl
42	49	100.0	694	2	US-08-007-999B-5	Sequence 5, Appl
43	49	100.0	694	2	US-08-689-276A-5	Sequence 5, Appl
44	49	100.0	695	1	US-08-371-930-27	Sequence 27, Appl
45	49	100.0	695	1	US-08-123-702-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
PCT-US94-07043A-7
; Sequence 7, Application PC/TUS9407043A
; GENERAL INFORMATION:
; APPLICANT: Tamburini, Paul P.; Benz, G nter; H bich,
; APPLICANT: Dieter; Dreyer, Robert N.; Koenig, Gerhard
; TITLE OF INVENTION: CATHEPSIN D IS AN ANVLOIDGENIC
; TITLE OF INVENTION: PROTEASE IN ALZHEIMER S DISEASE
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Miles Inc.
; STREET: 400 Morgan Lane
; CITY: West Haven
; STATE: Connecticut
; COUNTRY: USA
; ZIP: 06516
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 800 kb storage
; COMPUTER: Sharp PC 4500
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/07043A
; FILING DATE: June 21, 1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/10889
; FILING DATE: November 12, 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/995,660
; FILING DATE: December 16, 1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/880,914
; FILING DATE: May 11, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Pamela A. Simonton
; REGISTRATION NUMBER: 31,060
; REFERENCE/DOCKET NUMBER: MTI 224.3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (203) 937-2340
; TELEFAX: (203) 937-2795
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; PCT-US94-07043A-7

3

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/141,324
; FILING DATE: 21-OCT-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 44-93
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303-499-8080
; TELEFAX: 303-499-8089
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 27 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-141-324-11

Query Match 100.0%; Score 49; DB 1; Length 27;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVKMDAEFR 10
Db 4 SEVKMDAEFR 13

RESULT 5
US-08-541-902-11
; Sequence 11, Application US/08541902
; Patent No. 5707620
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S.
; APPLICANT: Barr, Philip J.
; APPLICANT: Pavloff, Nadine
; APPLICANT: Pike, Robert N.
; TITLE OF INVENTION: Lysine-specific Porphyromonas gingivalis
; TITLE OF INVENTION: Protease
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee and Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: CO
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/541,902
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/141,324
; FILING DATE: 21-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 44-93
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303-499-8080

; TELEFAX: 303-499-8089
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 27 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-541-902-11

Query Match 100.0%; Score 49; DB 1; Length 27;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVKMDAEFR 10
Db 4 SEVKMDAEFR 13

RESULT 6
US-08-462-859A-5
; Sequence 5, Application US/08462859A
; Patent No. 5652092
; GENERAL INFORMATION:
; APPLICANT: Jacobsen, J. S.
; APPLICANT: Vittek, M. P.
; TITLE OF INVENTION: No. 5652092el Amyloid Precursor and Method of
; TITLE OF INVENTION: Using Same to Access Agents Which Down-Regulate Formation
; TITLE OF INVENTION: of B-Amyloid Peptide
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Cyanamid Company
; STREET: One Cyanamid Plaza
; CITY: Wayne
; STATE: New Jersey
; COUNTRY: United States
; ZIP: 07470-8426
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,859A
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Barnhard, Elizabeth M.
; REGISTRATION NUMBER: 31,088
; REFERENCE/DOCKET NUMBER: 31,844-04
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201)831-3246
; TELEFAX: (201)831-3305
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 45 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-462-859A-5

Query Match 100.0%; Score 49; DB 1; Length 45;
Best Local Similarity 100.0%; Pred. No. 0.0017;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVKMDAEFR 10
Db 5 SEVKMDAEFR 14

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/464,247A
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Barnhard, Elizabeth M.
REGISTRATION NUMBER: 31,088
REFERENCE/DOCKET NUMBER: 31,844-03
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-683-2158
TELEFAX: 201-683-4117
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 45 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-464-247A-5

Query Match 100.0%; Score 49; DB 1; Length 45;
Best Local Similarity 100.0%; Pred. No. 0.0017;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVKMDAEFR 10
Db 5 SEVKMDAEFR 14

RESULT 9
US-08-464-248A-5
Sequence 5, Application US/08464248A
Patent No. 5703209
GENERAL INFORMATION:
APPLICANT: Jacobsen, J. S.
TITLE OF INVENTION: No. 5703209el Amyloid Precursor and Method of
TITLE OF INVENTION: Using Same to Access Agents Which Down-Regulate Formation
TITLE OF INVENTION: of B-Amyloid Peptide
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: American Cyanamid Company
STREET: One Cyanamid Plaza
CITY: Wayne
STATE: New Jersey
COUNTRY: United States
ZIP: 07470-8426
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/464,248A
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Barnhard, Elizabeth M.
REGISTRATION NUMBER: 31,088
REFERENCE/DOCKET NUMBER: 31,844-02
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201)831-3246
TELEFAX: (201)831-3305
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 45 amino acids
TYPE: amino acid
STRANDEDNESS:

US-08-123-659A-5
Sequence 5, Application US/08123659A
Patent No. 5656477
GENERAL INFORMATION:
APPLICANT: Jacobsen, J. S.
TITLE OF INVENTION: No. 5656477el Amyloid Precursor and Method of
TITLE OF INVENTION: Using Same to Access Agents Which Down-Regulate Formation
TITLE OF INVENTION: of B-Amyloid Peptide
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Anne Rosenblum
STREET: 163 Delaware Avenue, Suite 212
CITY: Delmar
STATE: New York
COUNTRY: U.S.A.
ZIP: 12054
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/123,659A
FILING DATE: 20-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rosenblum, Anne M.
REGISTRATION NUMBER: 30,419
REFERENCE/DOCKET NUMBER: 31,844-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (518)475-0611
TELEFAX: (518)475-0619
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 45 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-123-659A-5

Query Match 100.0%; Score 49; DB 1; Length 45;
Best Local Similarity 100.0%; Pred. No. 0.0017;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVKMDAEFR 10
Db 5 SEVKMDAEFR 14

RESULT 8
US-08-464-247A-5
Sequence 5, Application US/08464247A
Patent No. 5693478
GENERAL INFORMATION:
APPLICANT: Jacobsen, J. S.
TITLE OF INVENTION: No. 5693478el Amyloid Precursor and Method of
TITLE OF INVENTION: Using Same to Access Agents Which Down-Regulate Formation
TITLE OF INVENTION: of B-Amyloid Peptide
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: American Cyanamid Company
STREET: One Campus Drive
CITY: Parsippany
STATE: New Jersey
COUNTRY: United States
ZIP: 07054
COMPUTER READABLE FORM:

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-464-248A-5

Query Match 100.0%; Score 49; DB 1; Length 45;
Best Local Similarity 100.0%; Pred. No. 0.0017;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SEVKMDAEFR 10
Db 5 SEVKMDAEFR 14

RESULT 10

US-08-371-930-25
; Sequence 25, Application US/08371930
; Patent No. 5578451
; GENERAL INFORMATION:
; APPLICANT: Nishimoto, Ikuro
; TITLE OF INVENTION: ALZHEIMER'S DISEASE THERAPEUTICS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
FILING DATE:
PRIORITY APPLICATION NUMBER: US/08/371,930
CLASSIFICATION: 436
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/019,208
FILING DATE: February 18, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/154001
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 58
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear

PCT-US94-01712-25

Query Match 100.0%; Score 49; DB 1; Length 58;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SEVKMDAEFR 10
Db 42 SEVKMDAEFR 51

Query Match 100.0%; Score 49; DB 1; Length 58;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SEVKMDAEFR 10
Db 42 SEVKMDAEFR 51

RESULT 11

PCT-US94-01712-25
; Sequence 25, Application PC/TUS9401712
; GENERAL INFORMATION:
; APPLICANT: Nishimoto, Ikuro
; TITLE OF INVENTION: ALZHEIMER'S DISEASE THERAPEUTICS

NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01712
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/019,208
FILING DATE: February 18, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/154001
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 58
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear

PCT-US94-01712-25

Query Match 100.0%; Score 49; DB 5; Length 58;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SEVKMDAEFR 10
Db 42 SEVKMDAEFR 51

RESULT 12

US-08-462-859A-3
; Sequence 3, Application US/08462859A
; Patent No. 5632092
; GENERAL INFORMATION:
; APPLICANT: Jacobsen, J. S.
; APPLICANT: Vitek, M. P.
; TITLE OF INVENTION: No. 5652092el Amyloid Precursor and Method of
; TITLE OF INVENTION: Using Same to Access Agents Which Down-Regulate Formation
; TITLE OF INVENTION: of B-Amyloid Peptide
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Cyanamid Company
; STREET: One Cyanamid Plaza
; CITY: Wayne
; STATE: New Jersey
; COUNTRY: United States
; ZIP: 07470-8426
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,859A
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435

Query Match 100.0%; Score 49; DB 1; Length 58;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SEVKMDAEFR 10
Db 42 SEVKMDAEFR 51

ATTORNEY/AGENT INFORMATION:
NAME: Barnhard, Elizabeth M.
REGISTRATION NUMBER: 31,088
REFERENCE/DOCKET NUMBER: 31,844-04
TELEPHONE: (201)831-3246
TELEFAX: (201)831-3305
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 63 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-462-859A-3

Query Match 100.0%; Score 49; DB 1; Length 63;
Best Local Similarity 100.0%; Pred. No. 0.0025;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVKMDAEFR 10
Db 5 SEVKMDAEFR 14

RESULT 13
US-08-462-859A-4
Sequence 4, Application US/08462859A
Patent No. 5652092
GENERAL INFORMATION:
APPLICANT: Jacobsen, J. S.
APPLICANT: Vitek, M. P.
TITLE OF INVENTION: No. 5652092el Amyloid Precursor and Method of
TITLE OF INVENTION: Using Same to Access Agents Which Down-Regulate Formation
TITLE OF INVENTION: of B-Amyloid Peptide
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: American Cyanamid Company
STREET: One Cyanamid Plaza
CITY: Wayne
STATE: New Jersey
COUNTRY: United States
ZIP: 07470-8426
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,859A
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Barnhard, Elizabeth M.
REGISTRATION NUMBER: 31,088
REFERENCE/DOCKET NUMBER: 31,844-04
TELEPHONE: (201)831-3246
TELEFAX: (201)831-3305
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 63 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-462-859A-4

Query Match 100.0%; Score 49; DB 1; Length 63;
Best Local Similarity 100.0%; Pred. No. 0.0025;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVKMDAEFR 10
Db 5 SEVKMDAEFR 14

RESULT 14
US-08-123-659A-3
Sequence 3, Application US/08123659A
Patent No. 5656477
GENERAL INFORMATION:
APPLICANT: Jacobsen, J. S.
APPLICANT: Vitek, M. P.
TITLE OF INVENTION: No. 5656477el Amyloid Precursor and Method of
TITLE OF INVENTION: Using Same to Access Agents Which Down-Regulate Formation
TITLE OF INVENTION: of B-Amyloid Peptide
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Anne Rosenblum
STREET: 163 Delaware Avenue, Suite 212
CITY: Delmar
STATE: New York
COUNTRY: U.S.A.
ZIP: 12054
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/123,659A
FILING DATE: 20-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rosenblum, Anne M.
REGISTRATION NUMBER: 30,419
REFERENCE/DOCKET NUMBER: 31,844-01
TELEPHONE: (518)475-0611
TELEFAX: (518)475-0619
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 63 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-123-659A-3

Query Match 100.0%; Score 49; DB 1; Length 63;
Best Local Similarity 100.0%; Pred. No. 0.0025;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

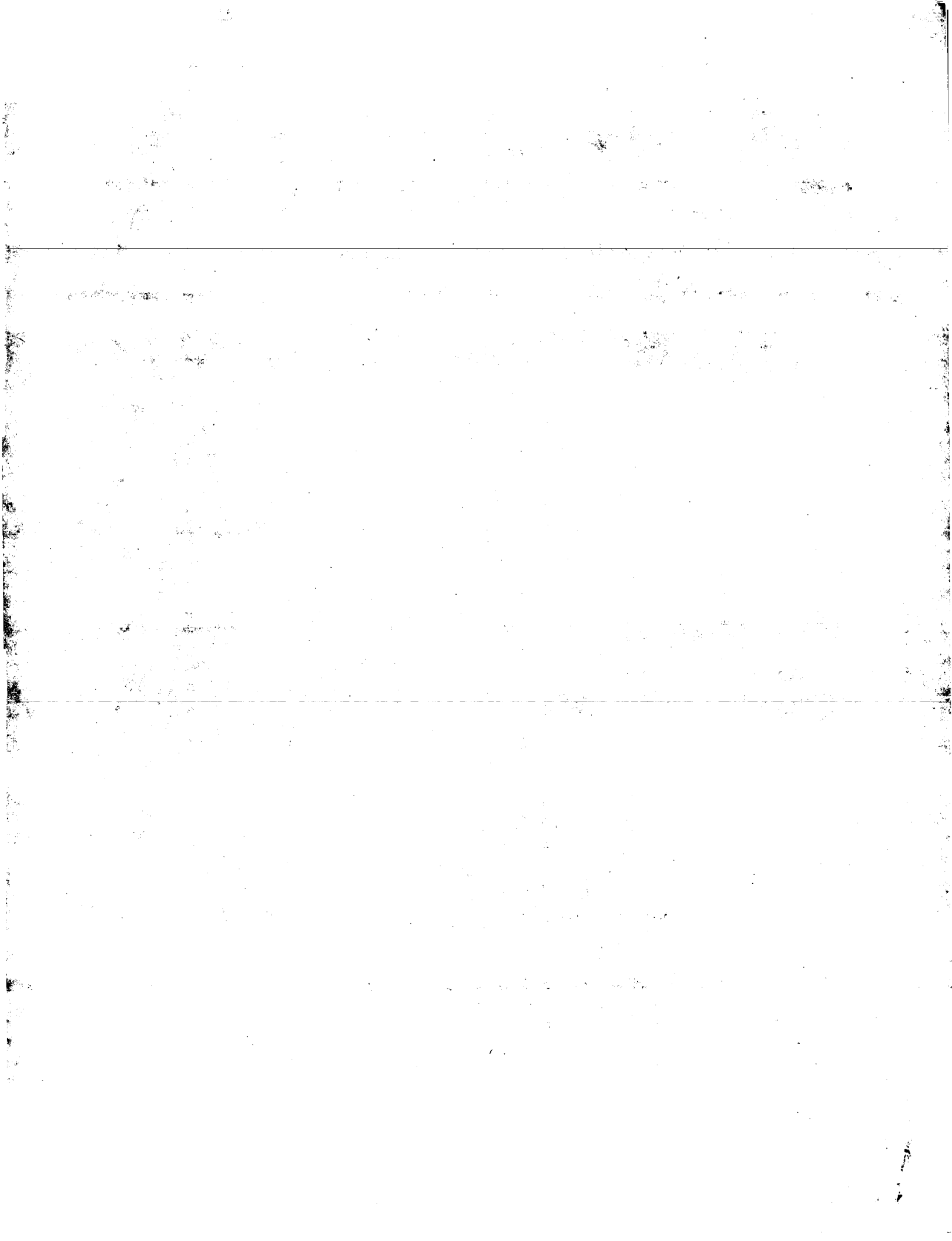
QY 1 SEVKMDAEFR 10
Db 5 SEVKMDAEFR 14

RESULT 15
US-08-123-659A-4
Sequence 4, Application US/08123659A
Patent No. 5656477
GENERAL INFORMATION:
APPLICANT: Jacobsen, J. S.
APPLICANT: Vitek, M. P.
TITLE OF INVENTION: No. 5656477el Amyloid Precursor and Method of
TITLE OF INVENTION: Using Same to Access Agents Which Down-Regulate Formation
TITLE OF INVENTION: of B-Amyloid Peptide
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Anne Rosenblum
STREET: 163 Delaware Avenue, Suite 212
CITY: Delmar

STATE: New York
COUNTRY: U.S.A.
ZIP: 12054
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/123,659A
FILING DATE: 20-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rosenblum, Anne M.
REGISTRATION NUMBER: 30,419
REFERENCE/DOCKET NUMBER: 31,844-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (518)475-0611
TELEFAX: (518)475-0619
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 63 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-123-659A-4

Query Match 100.0%; Score 49; DB 1; Length 63;
Best Local Similarity 100.0%; Pred. No. 0.0025;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 SEVKMDAEFR 10
Db 5 SEVKMDAEFR 14

Search completed: September 6, 2001, 16:39:35
Job time: 124 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:45:46 ; Search time 134.15 Seconds
(without alignments)
5.678 Million cell updates/sec

Title: US-09-603-713-4

Perfect score: 49

Sequence: 1 SEVKMDAEFR 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_68.*

1: p1r1.*

2: p1r2.*

3: p1r3.*

4: p1r4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	49	100.0	57	2 E60045	Alzheimer's diseases
2	49	100.0	57	2 F60045	Alzheimer's diseases
3	49	100.0	57	2 G60045	Alzheimer's diseases
4	49	100.0	57	2 D60045	Alzheimer's diseases
5	49	100.0	57	2 A60045	Alzheimer's diseases
6	49	100.0	57	2 B60045	Alzheimer's diseases
7	49	100.0	82	2 P00438	Alzheimer's diseases
8	49	100.0	695	1 A49795	Alzheimer's diseases
9	49	100.0	770	1 Q8H0A4	Alzheimer's diseases
10	44	89.8	33	2 S23094	beta-amyloid prote
11	44	89.8	695	2 A27485	Alzheimer's diseases
12	44	89.8	695	2 S00550	Alzheimer's diseases
13	43	87.8	747	2 JH0773	Alzheimer's diseases
14	34	69.4	3562	2 A47171	chondroitin sulfat
15	34	69.4	4563	1 LPHUB	apolipoprotein B-1
16	33	67.3	1245	2 T38127	phosphoprotein - f
17	33	67.3	1245	2 G86404	probable P-glycop
18	32	65.3	263	2 D84226	hypothetical prote
19	32	65.3	354	2 S51143	FMO-protein - Chlo
20	32	65.3	392	2 T49471	mucin (muc3) relat
21	32	65.3	426	2 G75187	probable trehalose
22	32	65.3	625	2 D86244	protein Ser/Thr pr
23	32	65.3	700	2 E84131	transcription anti
24	32	65.3	929	2 T52517	hypothetical prote
25	32	65.3	1265	2 T51498	hypothetical prote
26	32	65.3	2514	2 T37320	ataxia telangiecta
27	32	65.3	2619	2 T24588	hypothetical prote
28	31	63.3	155	2 F75040	hypothetical prote
29	31	63.3	178	2 C64168	hypothetical prote

30 31 63.3 183 2 S56460 hypothetical 21.4K
31 31 63.3 183 2 C86121 probable alpha hel
32 31 63.3 198 2 S48290 OX40 ligand - mous
33 31 63.3 199 2 F72060 conserved hypothet
34 31 63.3 199 2 C86564 CT471 hypothetical
35 31 63.3 226 2 G59129 hypothetical prote
36 31 63.3 279 2 T41124 single-stranded DN
37 31 63.3 282 2 T26112 hypothetical prote
38 31 63.3 286 2 G85230 hypothetical prote
39 31 63.3 389 2 G84245 NADH dehydrogenase
40 31 63.3 470 2 C75591 threonine synthase
41 31 63.3 479 2 S73921 MG098 homolog G07
42 31 63.3 626 2 T10237 RNA helicase RH16
43 31 63.3 692 2 E96841 hypothetical prote
44 31 63.3 1024 2 C64208 hypothetical prote
45 31 63.3 1398 2 H71606 hypothetical prote

ALIGNMENTS

RESULT 1

E60045
Alzheimer's disease amyloid beta/A4 protein precursor - sheep (fragment)
C:Species: Ovis sp. (sheep)
C:Date: 01-Dec-1992 #sequence_revision 01-Dec-1992 #text_change 28-Jul-1995
C:Accession: E60045
R:Johnstone, E.M.; Chaney, M.O.; Norris, F.H.; Pascual, R.; Little, S.P.
Brain Res. Mol. Brain Res. 10, 299-305, 1991
A:Title: Conservation of the sequence of the Alzheimer's disease amyloid peptide in d
A:Reference number: A60045; MUID:92017079
A:Accession: E60045
A:Molecule type: mRNA
A:Residues: 1-57 <JOH>
A:Cross-references: EMBL:X56130
C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinas
C:Keywords: alternative splicing; Alzheimer's disease; amyloid; brain

Query Match 100.0%; Score 49; DB 2; Length 57;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SEVKMDAEFR 10
Db 1 SEVKMDAEFR 10
|||||

RESULT 2

F60045
Alzheimer's disease amyloid beta/A4 protein precursor - pig (fragment)
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 01-Dec-1992 #sequence_revision 01-Dec-1992 #text_change 13-Aug-1999
C:Accession: F60045
R:Johnstone, E.M.; Chaney, M.O.; Norris, F.H.; Pascual, R.; Little, S.P.
Brain Res. Mol. Brain Res. 10, 299-305, 1991
A:Title: Conservation of the sequence of the Alzheimer's disease amyloid peptide in d
A:Reference number: A60045; MUID:92017079
A:Accession: F60045
A:Molecule type: mRNA
A:Residues: 1-57 <JOH>
A:Cross-references: EMBL:X56127; NID:gl1895; PIDN:CAA39592.1; PID:gl1896
C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinas
C:Keywords: alternative splicing; Alzheimer's disease; amyloid; brain

Query Match 100.0%; Score 49; DB 2; Length 57;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SEVKMDAEFR 10
Db 1 SEVKMDAEFR 10
|||||

RESULT 3
 G60045
 Alzheimer's disease amyloid beta/A4 protein precursor - guinea pig (fragment)
 C:Species: Cavia porcellus (guinea pig)
 C:Date: 01-Dec-1992 #sequence_revision 01-Dec-1992 #text_change 28-Jul-1995
 C:Accession: G60045
 R:Johnstone, E.M.; Chaney, M.O.; Norris, F.H.; Pascual, R.; Little, S.P.
 Brain Res. Mol. Brain Res. 10, 299-305, 1991
 A:Title: Conservation of the sequence of the Alzheimer's disease amyloid peptide in dog,
 A:Reference number: A60045; MUID:92017079
 A:Accession: G60045
 A:Molecule type: mRNA
 A:Residues: 1-57 <JOH>
 A:CROSS-references: EMBL:X56126
 C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinase
 C:Keywords: alternative splicing; Alzheimer's disease; amyloid; brain

 Query Match 100.0%; Score 49; DB 2; Length 57;
 Best Local Similarity 100.0%; Pred. No. 0.0013;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 1 SEVKMDAEFR 10
 |||||
 DB 1 SEVKMDAEFR 10

 RESULT 4
 D60045
 Alzheimer's disease amyloid beta/A4 protein precursor - bovine (fragment)
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 01-Dec-1992 #sequence_revision 01-Dec-1992 #text_change 28-Jul-1995
 C:Accession: D60045
 R:Johnstone, E.M.; Chaney, M.O.; Norris, F.H.; Pascual, R.; Little, S.P.
 Brain Res. Mol. Brain Res. 10, 299-305, 1991
 A:Title: Conservation of the sequence of the Alzheimer's disease amyloid peptide in dog,
 A:Reference number: A60045; MUID:92017079
 A:Accession: D60045
 A:Molecule type: mRNA
 A:Residues: 1-57 <JOH>
 A:CROSS-references: EMBL:X56124
 C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinase
 C:Keywords: alternative splicing; Alzheimer's disease; amyloid; brain

 Query Match 100.0%; Score 49; DB 2; Length 57;
 Best Local Similarity 100.0%; Pred. No. 0.0013;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 1 SEVKMDAEFR 10
 |||||
 DB 1 SEVKMDAEFR 10

 RESULT 5
 A60045
 Alzheimer's disease amyloid beta/A4 protein precursor - dog (fragment)
 C:Species: Canis lupus familiaris (dog)
 C:Date: 01-Dec-1992 #sequence_revision 01-Dec-1992 #text_change 28-Jul-1995
 C:Accession: A60045
 R:Johnstone, E.M.; Chaney, M.O.; Norris, F.H.; Pascual, R.; Little, S.P.
 Brain Res. Mol. Brain Res. 10, 299-305, 1991
 A:Title: Conservation of the sequence of the Alzheimer's disease amyloid peptide in dog,
 A:Reference number: A60045; MUID:92017079
 A:Accession: A60045
 A:Molecule type: mRNA
 A:Residues: 1-57 <JOH>
 A:CROSS-references: EMBL:X56125
 C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinase
 C:Keywords: alternative splicing; Alzheimer's disease; amyloid; brain

Query Match 100.0%; Score 49; DB 2; Length 57;
 Best Local Similarity 100.0%; Pred. No. 0.0013;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 1 SEVKMDAEFR 10
 |||||
 DB 1 SEVKMDAEFR 10

 RESULT 6
 B60045
 Alzheimer's disease amyloid beta/A4 protein precursor - polar bear (fragment)
 C:Species: Ursus maritimus (polar bear)
 C:Date: 01-Dec-1992 #sequence_revision 01-Dec-1992 #text_change 13-Aug-1999
 C:Accession: B60045
 R:Johnstone, E.M.; Chaney, M.O.; Norris, F.H.; Pascual, R.; Little, S.P.
 Brain Res. Mol. Brain Res. 10, 299-305, 1991
 A:Title: Conservation of the sequence of the Alzheimer's disease amyloid peptide in d
 A:Reference number: A60045; MUID:92017079
 A:Accession: B60045
 A:Molecule type: mRNA
 A:Residues: 1-57 <JOH>
 A:CROSS-references: EMBL:X56128; NID:92165; PIDN:CAA39593.1; PID:g2166
 C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinase
 C:Keywords: alternative splicing; Alzheimer's disease; amyloid; brain

 Query Match 100.0%; Score 49; DB 2; Length 57;
 Best Local Similarity 100.0%; Pred. No. 0.0013;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 1 SEVKMDAEFR 10
 |||||
 DB 1 SEVKMDAEFR 10

 RESULT 7
 P00438
 Alzheimer's disease amyloid A4 protein precursor - rabbit (fragment)
 C:Species: Oryctolagus cuniculus (domestic rabbit)
 C:Date: 30-Sep-1993 #sequence_revision 19-Oct-1995 #text_change 19-Oct-1995
 C:Accession: P00438; C60045
 R:Davidson, J.S.; West, R.L.; Kotikalapudi, P.; Maroun, L.E.
 Biochem. Biophys. Res. Commun. 188, 905-911, 1992
 A:Title: Sequence and methylation in the beta/A4 region of the rabbit amyloid precurs
 A:Reference number: P00438; MUID:93075180
 A:Accession: P00438
 A:Molecule type: DNA
 A:Residues: 1-82 <DAV>
 A:CROSS-references: GB:M83558; GB:M83657
 R:Johnstone, E.M.; Chaney, M.O.; Norris, F.H.; Pascual, R.; Little, S.P.
 Brain Res. Mol. Brain Res. 10, 299-305, 1991
 A:Title: Conservation of the sequence of the Alzheimer's disease amyloid peptide in d
 A:Reference number: A60045; MUID:92017079
 A:Accession: C60045
 A:Molecule type: mRNA
 A:Residues: 12-68 <JOH>
 A:CROSS-references: EMBL:X56129
 C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinase
 C:Keywords: alternative splicing; Alzheimer's disease; amyloid; Down's syndrome

 Query Match 100.0%; Score 49; DB 2; Length 82;
 Best Local Similarity 100.0%; Pred. No. 0.002;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 1 SEVKMDAEFR 10
 |||||
 DB 12 SEVKMDAEFR 21

 RESULT 8

A49795
 Alzheimer's disease amyloid beta protein precursor - crab-eating macaque
 C:Species: Macaca fascicularis (crab-eating macaque)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: A49795
 R:Podlinsky, M.B.; Tolan, D.R.; Selkoe, D.J.
 Am. J. Pathol. 138, 1423-1435, 1991
 A:Title: Homology of the amyloid beta protein precursor in monkey and human supports a beta-amyloid precursor hypothesis
 A:Reference number: A49795; MUID:91273117
 A:Accession: A49795
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-695 <POD>
 A:Cross-references: GB:M58727; NID:g342062; PIDN:AAA36829.1; PID:g342063
 C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinase 1
 C:Keywords: alternative splicing

Query Match 100.0%; Score 49; DB 1; Length 695;
 Best Local Similarity 100.0%; Pred. No. 0.021;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SEVKMDAEFR 10
 |||||
 Db 592 SEVKMDAEFR 601

RESULT 9
 ORHU44
 Alzheimer's disease amyloid beta protein precursor [validated] - human
 N:Alternate names: Alzheimer's disease amyloid A4 protein; coagulation factor XIA inhibitor
 N:Contains: amyloid beta protein long, plaque form; amyloid beta protein short, vascular
 protein precursor splice form APP(770)
 C:Species: Homo sapiens (man)
 C:Date: 30-Jun-1987 #sequence_revision 28-Jul-1995 #text_change 15-Sep-2000
 C:Accession: S02260; S05194; A32277; A32260; A35486; I39452; I39451; I39453; I59562; A44
 4688; A28583; A29302; A60805; J00038; S06121; A60355; A59011; A38384; S29076; S38252; S3
 R:Lemaire, H.G.; Salbaum, J.M.; Multhaup, G.; Kang, J.; Bayney, R.M.; Unterbeck, A.; Bey
 Nucleic Acids Res. 17, 517-522, 1989
 A:Title: The PreA4(695) precursor protein of Alzheimer's disease A4 amyloid is encoded b
 A:Reference number: S02260; MUID:89128427
 A:Accession: S02260
 A:Molecule type: DNA
 A:Residues: 1-288, 'V', 365-770 <LEM1>
 A:Cross-references: EMBL:X13466
 A:Note: alternative splice form APP(695)
 R:Lemaire, H.G.
 submitted to the EMBL Data Library, November 1988
 A:Reference number: S05194
 A:Accession: S05194
 A:Molecule type: DNA
 A:Residues: 1-14, 'VW', 17-288, 'V', 365-770 <LEM2>
 A:Cross-references: EMBL:X13466; NID:g355598; PIDN:CAA31830.1; PID:g871360
 A:Note: alternative splice form APP(695)
 R:La Fauci, G.; Lahiri, D.K.; Salton, S.R.J.; Robakis, N.K.
 Biochem. Biophys. Res. Commun. 159, 297-304, 1989
 A:Title: Characterization of the 5'-end region and the first two exons of the beta-prote
 A:Reference number: A32277; MUID:89165870
 A:Accession: A32277
 A:Molecule type: DNA
 A:Residues: 1-75 <LAF>
 A:Cross-references: GB:M24546; GB:M24547; NID:g341202; PIDN:AAC13654.1; PID:g516074
 R:Johnstone, E.M.; Chaney, M.O.; Moore, R.E.; Ward, K.E.; Norris, F.H.; Little, S.P.
 Biochem. Biophys. Res. Commun. 163, 1248-1255, 1989
 A:Title: Alzheimer's disease amyloid peptide is encoded by two exons and shows similarit
 A:Reference number: A33260; MUID:89392030
 A:Accession: A33260
 A:Molecule type: DNA
 A:Residues: 656-737
 A:Cross-references: GB:M29270; NID:g178863; PIDN:AAA51768.1; PID:g178865
 R:Prelli, F.; Levy, E.; van Duinen, S.G.; Bots, G.T.A.M.; Luyendijk, W.; Frangione, B.
 Biochem. Biophys. Res. Commun. 170, 301-307, 1990
 A:Title: Expression of a normal and variant Alzheimer's beta-protein gene in amyloid of

A:Reference number: A35486; MUID:90321244
 A:Accession: A35486
 A:Molecule type: DNA
 A:Residues: 672-710 <PRE1>
 A:Note: 693-Gln was found in DNA isolated from HCHWA-D patients
 R:Yoshikai, S.I.; Sasaki, H.; Doh-ura, K.; Furuya, H.; Sakaki, Y.
 Gene 87, 257-263, 1990
 A:Title: Genomic organization of the human amyloid beta-protein precursor gene.
 A:Reference number: I39451; MUID:90236318
 A:Accession: I39452
 A>Status: nucleic acid sequence not shown; translation not shown; translated from GB/
 A:Molecule type: DNA
 A:Residues: 1-770 <YOS1>
 A:Cross-references: GB:M33112; NID:g178613; PIDN:AAB59502.1; PID:g178616
 A:Accession: I39451
 A>Status: nucleic acid sequence not shown; translation not shown; translated from GB/
 A:Molecule type: DNA
 A:Residues: 1-530, 'QWLMPVIPAFWEAKVGR' <YOS2>
 A:Cross-references: GB:M34875; NID:g178608; PIDN:AAB59501.1; PID:g178615
 R:Yoshikai, S.I.; Sasaki, H.; Doh-ura, K.; Furuya, H.; Sakaki, Y.
 Gene 102, 291-292, 1991
 A:Reference number: A59020; MUID:91340168
 A:Contents: annotation; erratum
 A:Note: revised physical map for reference I39451
 R:Levy, E.; Carman, M.D.; Fernandez-Madrid, I.J.; Power, M.D.; Lieberburg, I.; van Du
 Science 248, 1124-1126, 1990
 A:Title: Mutation of the Alzheimer's disease amyloid gene in hereditary cerebral hemo
 A:Reference number: I39453; MUID:90260663
 A:Accession: I39453
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 656-737 <LEV>
 A:Cross-references: GB:M37896; NID:g178618; PIDN:AAA51727.1; PID:g178620
 A:Note: a mutation with 693-Gln is presented
 R:Murrell, J.; Farlow, M.; Ghetti, B.; Benson, M.D.
 Science 254, 97-99, 1991
 A:Title: A mutation in the amyloid precursor protein associated with hereditary Alzhe
 A:Reference number: I59562; MUID:92022553
 A:Accession: I59562
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 689-716, 'P', 718-737 <MUR>
 A:Cross-references: GB:S57665; NID:g236720; PIDN:AAB19991.1; PID:g236721
 R:Kamino, K.; Orr, H.T.; Payami, H.; Wiljman, E.M.; Alonso, M.E.; Pulst, S.M.; Anders
 arakis, S.E.; Korenberg, J.R.; Sharma, V.; Kukull, W.; Larson, E.; Heston, L.L.; Mart
 Am. J. Hum. Genet. 51, 998-1014, 1992
 A:Title: Linkage and mutational analysis of familial Alzheimer disease kindreds for t
 A:Reference number: A44017; MUID:93035397
 A:Accession: A44017
 A:Molecule type: DNA
 A:Residues: 687-692, 'G', 694-718 <KAM1>
 A:Cross-references: GB:S45135; NID:g257377; PIDN:AAB23645.1; PID:g257378
 A:Experimental source: familial Alzheimer disease family SB
 A:Note: sequence extracted from NCBI backbone (NCBIP:115374)
 A:Accession: B44017
 A:Molecule type: DNA
 A:Residues: 687-718 <KAM2>
 A:Cross-references: GB:S45136; NID:g257379; PIDN:AAB23646.1; PID:g257380
 A:Experimental source: familial Alzheimer disease family LIT
 A:Note: sequence extracted from NCBI backbone (NCBIP:115376)
 A:Note: this sequence has a silent mutation
 R:Kang, J.; Lemaire, H.G.; Unterbeck, A.; Salbaum, J.M.; Masters, C.L.; Grzeschik, K.
 Nature 325, 733-736, 1987
 A:Title: The precursor of Alzheimer's disease amyloid A4 protein resembles a cell-sur
 A:Reference number: A03134; MUID:87144572
 A:Accession: A03134
 A:Molecule type: mRNA
 A:Residues: 1-288, 'V', 365-770 <KAN>
 A:Cross-references: GB:Y00264; NID:g28525; PIDN:CAA68374.1; PID:g28526
 A:Note: alternative splice form APP(695)
 R:Robakis, N.K.; Ramakrishna, N.; Wolfe, G.; Wisniewski, H.M.
 Proc. Natl. Acad. Sci. U.S.A. 84, 4190-4194, 1987
 A:Title: Molecular cloning and characterization of a cDNA encoding the cerebrovascula

A:Reference number: A29030; MUID:87231971
A:Accession: A29030
A:Molecule type: mRNA
A:Residues: 284-288, 'V', 365-646, 'E', 648-770 <ROB>
A:Cross-references: GB:M16765; NID:g178539; PIDN:AAA51722.1; PID:g178540
A:Note: the authors translated the codon GAG for residue 647 as Asp
R:Goldgaber, D.; Lerman, M.I.; McBride, O.W.; Saffioti, U.; Gajdusek, D.C.
Science 235, 877-880, 1987
A:Title: Characterization and chromosomal localization of a cDNA encoding brain amyloid
A:Reference number: A47584; MUID:87120328
A:Accession: A47584
A:Molecule type: mRNA
A:Residues: 674-756, 'S', 758-770 <GOL>
A:Cross-references: GB:M15533; NID:g178706; PIDN:AAA35540.1; PID:g178707
A:Experimental source: brain
R:Tanzi, R.E.; Gusella, J.F.; Watkins, P.C.; Bruns, G.A.P.; St George-Hyslop, P.; Van Ke
Science 235, 880-884, 1987
A:Title: Amyloid beta protein gene: cDNA, mRNA distribution, and genetic linkage near th
A:Reference number: A47585; MUID:87120329
A:Accession: A47585
A:Molecule type: mRNA
A:Residues: 672-678 <YR>
R:Tanzi, R.E.; McClatchey, A.I.; Lampertl, E.D.; Villa-Komaroff, L.; Gusella, J.F.; Neve
Nature 331, 528-530, 1988
A:Title: Protease inhibitor domain encoded by an amyloid protein precursor mRNA associat
A:Reference number: S00707; MUID:88122640
A:Accession: S00707
A:Molecule type: mRNA
A:Residues: 286-344, 'I', 365-366 <TAN2>
A:Cross-references: EMBL:X06982; NID:g28817; PIDN:CAA30042.1; PID:g929612
A:Experimental source: promyelocytic leukemia cell line HL60
A:Note: alternative splice form APP(751)
R:Ponte, P.; Gonzalez-DeWhitt, P.; Schilling, J.; Miller, J.; Hsu, D.; Greenberg, B.; Da
Nature 331, 525-527, 1988
A:Title: A new A4 amyloid mRNA contains a domain homologous to serine proteinase inhibi
A:Reference number: S00925; MUID:88122639
A:Accession: S00925
A:Molecule type: mRNA
A:Residues: 1-344, 'I', 365-770 <PO2>
A:Cross-references: GB:X06989; EMBL:X00297; NID:g28720; PIDN:CAA30050.1; PID:g28721
A:Note: alternative splice form APP(751)
R:Kitaguchi, N.; Takahashi, Y.; Tokushima, Y.; Shiojiri, S.; Ito, H.
Nature 331, 530-532, 1988
A:Title: Novel precursor of Alzheimer's disease amyloid protein shows protease inhibitor
A:Reference number: A38949; MUID:88122641
A:Accession: A38949
A:Molecule type: mRNA
A:Residues: 287-367 <KIT>
A:Cross-references: GB:X06981; NID:g28816; PIDN:CAA30041.1; PID:g929611
A:Experimental source: glioblastoma cell line
A:Note: alternative splice form APP(770)
R:Vitek, M.P.; Rasool, C.G.; de Sauvage, F.; Vitek, S.M.; Bartus, R.T.; Beer, B.; Ashton
Brain Res. Mol. Brain Res. 4, 121-131, 1988
A:Title: Absence of mutation in the beta-amyloid cDNAs cloned from the brains of three E
A:Reference number: A30320
A:Accession: A30320
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 284-288, 'V', 365-770 <VIT1>
A:Accession: B30320
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 122-288, 'V', 365-770 <VIT2>
A:Accession: C30320
A>Status: not compared with conceptual translation

A:Molecule type: mRNA
A:Residues: 608-770 <VIT3>
R:Zain, S.B.; Salim, M.; Chou, W.G.; Sajdel-Sulkowska, E.M.; Majocha, R.E.; Marotta,
Proc. Natl. Acad. Sci. U.S.A. 85, 929-933, 1988
A:Title: Molecular cloning of amyloid cDNA derived from mRNA of the Alzheimer disease
A:Reference number: A31087; MUID:88124954
A:Accession: A31087
A:Molecule type: mRNA
A:Residues: 507-770 <ZAI>
A:Cross-references: GB:M18734; NID:g178572; PIDN:AAA51726.1; PID:g178573
A:Note: the authors translated the codon GAA for residue 599 as Gly, ACC for residue
8 as Val, GTG for residue 609 as Asn, AAT for residue 610 as Gly, and GGT for residue
A:Note: the cited Genbank accession number, J03594, is not in release 101.0
R:Masters, C.L.; Multhaup, G.; Simms, G.; Pottgiesser, J.; Martins, R.N.; Beyreuther,
Query Match 100.0%; Score 49; DB 1; Length 770;
Best Local Similarity 100.0%; Pred. No. 0.024; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 0;
Qy 1 SEVKMDAEFR 10
Db 667 SEVKMDAEFR 676
|||||
RESULT 10
S23094
A:Title: beta-amyloid protein precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 03-May-1996
C:Accession: S23094
R:Kojima, S.; Omori, M.
FEBS Lett. 304, 57-60, 1992
A:Title: Two-way cleavage of beta-amyloid protein precursor by multicatalytic protein
A:Reference number: S23094; MUID:92316198
A:Accession: S23094
A:Molecule type: protein
A:Residues: 1-33 <KOU>
C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinas
Query Match 89.8%; Score 44; DB 2; Length 33;
Best Local Similarity 100.0%; Pred. No. 0.0085; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;
Qy 1 SEVKMDAEFR 9
Db 1 SEVKMDAEFR 9
|||||
RESULT 11
A27485
A:Title: Alzheimer's disease amyloid beta/A4 protein homolog precursor - mouse
N:Alternate names: proteinase nexin II
C:Species: Mus musculus (house mouse)
C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 13-Aug-1999
C:Accession: A27485; S19727; I49485
R:Yanada, T.; Sakaki, H.; Furuya, H.; Miyata, T.; Goto, I.; Sakaki, Y.
Biochem. Biophys. Res. Commun. 149, 665-671, 1987
A:Title: Complementary DNA for the mouse homolog of the human amyloid beta protein pr
A:Reference number: A27485; MUID:88106489
A:Accession: A27485
A:Molecule type: mRNA
A:Residues: 1-695 <YAM>
A:Cross-references: GB:M18373; NID:g191568; PIDN:AAA37139.1; PID:g309085
A:Experimental source: brain
R:de Strooper, B.; van Leuven, F.; van den Berghe, H.
Biochim. Biophys. Acta 1129, 141-143, 1991
A:Title: The amyloid beta protein precursor or proteinase nexin II from mouse is clos
A:Reference number: S19727; MUID:92096458
A:Accession: S19727
A:Molecule type: mRNA
A:Residues: 1-210, 'G', 212-220, 'S', 222-396, 'A', 398-402, 'T', 404-448, 'A', 450-695 <STR>
A:Cross-references: EMBL:X59379

R:Izum, R.; Yamada, T.; Yoshikai, S.; Sasaki, H.; Hattori, M.; Sakaki, Y.
Gene 112, 189-195, 1992
A:Title: Positive and negative regulatory elements for the expression of the Alzheimer's
A:Reference number: 149485; MUID:92209998
A:Accession: 149485
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-19 <RES>
A:Cross-references: GB:D10603; NID:g220328; PIDN:BAA01456.1; PID:g220329
C:Genetics:
A:Map position: 16C3
C:Superfamily: Alzheimer's disease amyloid beta protein; animal kunitz-type proteinase
C:Keywords: alternative splicing; amyloid; transmembrane protein

Query Match 89.8%; Score 44; DB 2; Length 695;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVKMDAEF 9
|||||
DB 592 SEVKMDAEF 600

RESULT 12
S00550
Alzheimer's disease amyloid beta protein precursor - rat
N:Alternate names: beta-A4 amyloid protein
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 13-Aug-1999
C:Accession: S00550; A41245; A39820; S46251
R:Shivers, B.D.; Hilbich, C.; Multhaup, G.; Salbaum, M.; Beyreuther, K.; Seeburg, P.H.
EMBO J. 7, 1365-1370, 1988
A:Title: Alzheimer's disease amyloidogenic glycoprotein: expression pattern in rat brain
A:Reference number: S00550; MUID:88312583
A:Accession: S00550
A:Molecule type: mRNA
A:Residues: 1-695 <SH>
A:Cross-references: EMBL:X07648; NID:g55616; PIDN:CAA30488.1; PID:g55617
R:Schubert, D.; Schroeder, R.; LaCorbiere, M.; Saitoh, T.; Cole, G.
Science 241, 223-226, 1988
A:Title: Amyloid beta protein precursor is possibly a heparan sulfate proteoglycan core
A:Reference number: A41245; MUID:88264430
A:Accession: A41245
A:Molecule type: protein
A:Residues: 18-37, 'X', 39-40, 'X', 42-44 <SCH>
A:Note: evidence for heparan sulfate attachment
R:Hesse, L.; Behr, D.; Masters, C.L.; Multhaup, G.
FEBS Lett. 349, 109-116, 1994
A:Title: The beta-A4 amyloid precursor protein binding to copper.
A:Reference number: S46251; MUID:94320627
A:Contents: annotation; copper binding sites
A:Note: rat peptides were isolated but not sequenced
R:Potempska, A.; Styles, J.; Mehta, P.; Kim, K.S.; Miller, D.L.
J. Biol. Chem. 266, 8464-8469, 1991
A:Title: Purification and tissue level of the beta-amyloid peptide precursor of rat brain
A:Reference number: A39820; MUID:91217087
A:Accession: A39820
A>Status: preliminary
A:Molecule type: protein
A:Residues: 18-32 <POT>
A:Experimental source: brain
C:Comment: Deposition of amyloid protein as neurofibrillary tangles and/or plaques is ch
C:Superfamily: Alzheimer's disease amyloid beta protein; animal kunitz-type proteinase
C:Keywords: alternative splicing; amyloid; glycoprotein; transmembrane protein
F:625-648/Domain: transmembrane #status predicted <TM>

Query Match 89.8%; Score 44; DB 2; Length 695;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVKMDAEF 9

DB 592 SEVKMDAEF 600
|||||
RESULT 13
JH0773
Alzheimer's disease amyloid beta protein precursor - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 10-Jun-1993 #sequence_revision 10-Jun-1993 #text_change 13-Aug-1999
C:Accession: JH0773
R:Okado, H.; Okamoto, H.
Biochem. Biophys. Res. Commun. 189, 1561-1568, 1992
A:Title: A xenopus homologue of the human beta-amyloid precursor protein: development
A:Reference number: JH0773; MUID:93129227
A:Accession: JH0773
A:Molecule type: mRNA
A:Residues: 1-747 <OKA>
A:Cross-references: GB:S52417; NID:g263150; PIDN:AAB24853.1; PID:g263151
A:Experimental source: larva
C:Superfamily: Alzheimer's disease amyloid beta protein; animal kunitz-type proteinase
C:Keywords: alternative splicing; amyloid
F:287-337/Domain: animal kunitz-type proteinase inhibitor homology <BPI>

Query Match 87.8%; Score 43; DB 2; Length 747;
Best Local Similarity 80.0%; Pred. No. 0.44;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVKMDAEF 10
|||||
DB 644 SEVKMDSEYR 653

RESULT 14
A47171
chondroitin sulfate proteoglycan PG-M core protein - chicken
C:Species: Gallus gallus (chicken)
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
C:Accession: A47171
R:Shinomura, T.; Nishida, Y.; Ito, K.; Kimata, K.
J. Biol. Chem. 268, 14461-14469, 1993
A:Title: cDNA cloning of PG-M, a large chondroitin sulfate proteoglycan expressed dur
A:Reference number: A47171; MUID:93300846
A:Accession: A47171
A>Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-3562 <SH>
A:Cross-references: GB:D13542; NID:g391643; PIDN:BAA02742.1; PID:g391644
A:Experimental source: stage 22-23 developing limb buds
A:Note: sequence extracted from NCBI backbone (NCBI:134456, NCBI:134457)
C:Superfamily: chicken chondroitin sulfate proteoglycan PG-M core protein; c-type lec
F:166-243/Domain: link protein repeat homology <LNK1>
F:264-345/Domain: link protein repeat homology <LNK2>
F:3258-3289/Domain: EGF homology <EGF1>
F:3296-3327/Domain: EGF homology <EGF>
F:3334-3454/Domain: C-type lectin homology <LCH>
F:3461-3517/Domain: complement factor H repeat homology <FHD>

Query Match 69.4%; Score 34; DB 2; Length 3562;
Best Local Similarity 66.7%; Pred. No. 2.1e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEVKMDAEF 9
|:|:|
DB 1709 STIKLDAEF 1717

RESULT 15
LPHUB
apolipoprotein B-100 precursor - human
N:Contains: apolipoprotein B-26; apolipoprotein B-48; apolipoprotein B-74

A>Title: A partial cDNA clone for human apolipoprotein B.	
A:Reference number: A25774; MUID:85270450	
A:Accession: A25774	
A:Molecule type: mRNA	
A:Residues: 709-791,'SSSKAAASHGCPHSGD',810-906 <DE>	
A:Cross-references: GB:K03175; NID:g178821; PIDN:AAA51759.1; PID:g178822	
R:Carlsson, P.; Darnfors, C.; Olofsson, S.O.; Bjursell, G.	
Gene 49, 29-51, 1986	
A>Title: Analysis of the human apolipoprotein B gene; complete structure of the B-74	
A:Reference number: A91565; MUID:87191999	
A:Accession: A26533	
A:Molecule type: mRNA	
A:Residues: 1282-2721,2742-3290,'L',3292-3336,'N',3338-3948,'F',3950-3963,'Y',3965-41	
A:Cross-references: GB:M15421; NID:g178617; PIDN:AAA51758.1; PID:g178818	
R:Hardman, D.A.; Protter, A.A.; Chen, G.C.; Schilling, J.W.; Sato, K.Y.; Lau, K.; Yam	
Biochemistry 26, 5478-5486, 1987	
A>Title: Structural comparison of human apolipoproteins B-48 and B-100.	
A:Reference number: A29671; MUID:88050832	
A:Accession: A29671	
A:Molecule type: mRNA	
A:Residues: 1671-2323,'PYW',2327-2352,'H',2354-2398 <HAR>	
A:Cross-references: GB:M17367; NID:g178731; PIDN:AAA51741.1; PID:g178732	
R:Shoulders, C.C.; Myant, N.B.; Sidoli, A.; Rodriguez, J.C.; Cortese, C.; Baralle, F.	
Atherosclerosis 58, 277-289, 1985	
A>Title: Molecular cloning of human LDL apolipoprotein B cDNA. Evidence for more than	
A:Reference number: A90084; MUID:86130855	
A:Accession: A29287	
A:Molecule type: mRNA	
A:Residues: 3846-4298 <SHO>	
R:Pfitzner, R.; Wagener, R.; Stoffel, W.	
Biol. Chem. Hoppe-Seyler 367, 1077-1083, 1986	
A>Title: Isolation, expression and characterization of a human apolipoprotein B 100-s	
A:Reference number: A25572; MUID:87076044	
A:Accession: A25572	
A:Molecule type: mRNA	
A:Residues: 4219-4337,'S',4339-4563 <PFI>	
A:Cross-references: GB:M36676	
R:Wei, C.F.; Chen, S.H.; Yang, C.Y.; Marcel, Y.L.; Milne, R.W.; Li, W.H.; Sparrow, J.	
Proc. Natl. Acad. Sci. U.S.A. 82, 7265-7269; 1985	
A:Reference number: A24738; MUID:86042646	
A:Accession: A24738	
A:Molecule type: mRNA	
A:Residues: 'N',3729-3731,'I',3733-3875,'A',3877-3948,'F',3950-3963,'Y',3965-3982,'S',	
A:Cross-references: GB:M12413; NID:g178735; PIDN:AAA51742.1; PID:g178736	
R:Chen, S.H.; Habib, G.; Yang, C.Y.; Gu, Z.W.; Lee, B.R.; Weng, S.; Silberman, S.R.;	
Science 238, 363-366, 1987	
A>Title: Apolipoprotein B-48 is the product of a messenger RNA with an organ-specific	
A:Reference number: A40133; MUID:88018019	
A:Accession: B40133	
A:Molecule type: mRNA	
A:Residues: 2165-2179 <CHI>	
A:Cross-references: GB:M18036; NID:g178799; PIDN:AAA51754.1; PID:g178800	
A>Note: this mRNA includes the stop codon of the organ-specific mRNA for apoA8	
A:Accession: A40133	
A:Molecule type: protein	
A:Residues: 51-75;101-110;129-139;158-174;197-207;276-287;298-304;306-314;526-532;538-	
36;1486-1498;1537-1556;1563-1572;1601-1610;1647-1661;1697-1724;1770-1781;1859-1897;1919	
A>Note: these fragments were derived from apoA8	
R:Hardman, D.A.; Protter, A.A.; Schilling, J.W.; Kane, J.P.	
Biochem. Biophys. Res. Commun. 149, 1214-1219, 1987	
A>Title: Carboxyl terminal analysis of human B-48 protein confirms the novel mechanism	
A:Reference number: A28002; MUID:86106542	
A:Accession: A28002	
A:Molecule type: mRNA	
A:Residues: 2129-2179,2181-2235 <HA2>	
A:Cross-references: GB:M18471	
A:Experimental source: intestine	
A>Note: this mRNA from intestine includes a stop codon created by RNA editing in plac	
R:Mebrabian, M.; Schumaker, V.N.; Fareed, G.C.; West, R.; Johnson, D.F.; Kirchgessner	
Nucleic Acids Res. 13, 6937-6953, 1985	
A>Title: Human apolipoprotein B: identification of cDNA clones and characterization c	
A:Reference number: A24269; MUID:86041888	
A:Accession: A24269	

A:Molecule type: mRNA
A:Residues: 3056-3159 <MEH>
A:Cross-References: GB:X03045; NID:g28783; PIDN:CAA26850.1; PID:g929609
R:Hospatankar, A.V.; Higuchi, K.; Law, S.W.; Meglin, N.; Brewer Jr., H.B.
Biochem. Biophys. Res. Commun. 148, 279-285, 1987
A:Title: Identification of a novel in-frame translational stop codon in human intestine
A:Reference number: A29655; MUID:88049670
A:Accession: A29659
A:Molecule type: mRNA
A:Residues: 2169-2179 <HOS>
A:Note: the sequence shown represents the carboxyl end of apolipoprotein B-48
A:Note: two RNA species, 14.1kb and 7.5kb in length, were isolated from the human intestine
ch encodes the 250K apob-48, CAA encoding 2180-Gln is substituted by the stop codon TAA,
R:Yang, C.; Kim, T.W.; Weng, S.; Lee, B.; Yang, M.; Gotto Jr., A.M.
Proc. Natl. Acad. Sci. U.S.A. 87, 5523-5527, 1990
A:Title: Isolation and characterization of sulfhydryl and disulfide peptides of human ap
A:Reference number: A35783; MUID:90319144
A:Contents: disulfide bonds
A:Accession: A35783
A:Molecule type: Protein
A:Residues: 28-41; 76-97, 'I', 99-100; 175-193; 206-215; 239-249; 259-266; 357-399; 455-490; 512-5
A:Note: cysteines at positions 1112, 1422, 1505, 1662, 3761, 3917, and 4217 have free su
R:LeBoeuf, R.C.; Miller, C.; Shively, J.E.; Schumaker, V.N.; Balla, M.A.; Lusis, A.J.
FEBS Lett. 170, 105-108, 1984
A:Title: Human apolipoprotein B: partial amino acid sequence.
A:Reference number: A22006; MUID:84208786
A:Accession: A22006
A:Molecule type: Protein
A:Residues: 873-892, 'K', 894-896 <LE1>
A:Accession: B22006
A:Molecule type: Protein
A:Residues: 3113, 'L', 3115-3130, 'R', 3132-3133, 'P', 3135-3136, 'R' <LE2>
R:Blackhart, B.D.; Ludwig, E.M.; Pierotti, V.R.; Calati, L.; Onasch, M.A.; Wallis, S.C.;
J. Biol. Chem. 261, 15364-15367, 1986
A:Title: Structure of the human apolipoprotein B gene.
A:Reference number: A92564; MUID:87057153
A:Contents: annotation; gene structure
R:Wagener, R.; Pfitzner, R.; Stoffel, W.
Biol. Chem. Hoppe-Seyler 368, 419-425, 1987
A:Title: Studies on the organization of the human apolipoprotein B 100 gene.
A:Reference number: A90715; MUID:87271140
A:Contents: annotation; gene structure
R:Weisgraber, K.H.; Rall Jr., S.C.
J. Biol. Chem. 262, 11097-11103, 1987
A:Title: Human apolipoprotein B-100 heparin-binding sites.
A:Reference number: A92605; MUID:87280197
A:Contents: annotation; heparin binding and disulfide bond
R:Dashti, N.; Lee, D.M.; Mok, T.
Biochem. Biophys. Res. Commun. 137, 493-499, 1986
A:Title: Apolipoprotein B is a calcium binding protein.
A:Reference number: A90125; MUID:86242245
A:Contents: annotation; calcium binding
R:Carlsson, P.; Olofsson, S.O.; Bondjers, G.; Darnfors, C.; Wiklund, O.; Bjursell, G.
Nucleic Acids Res. 13, 8813-8826, 1985
A:Title: Molecular cloning of human apolipoprotein B cDNA.
A:Reference number: I37178; MUID:86093680
A:Accession: I37180

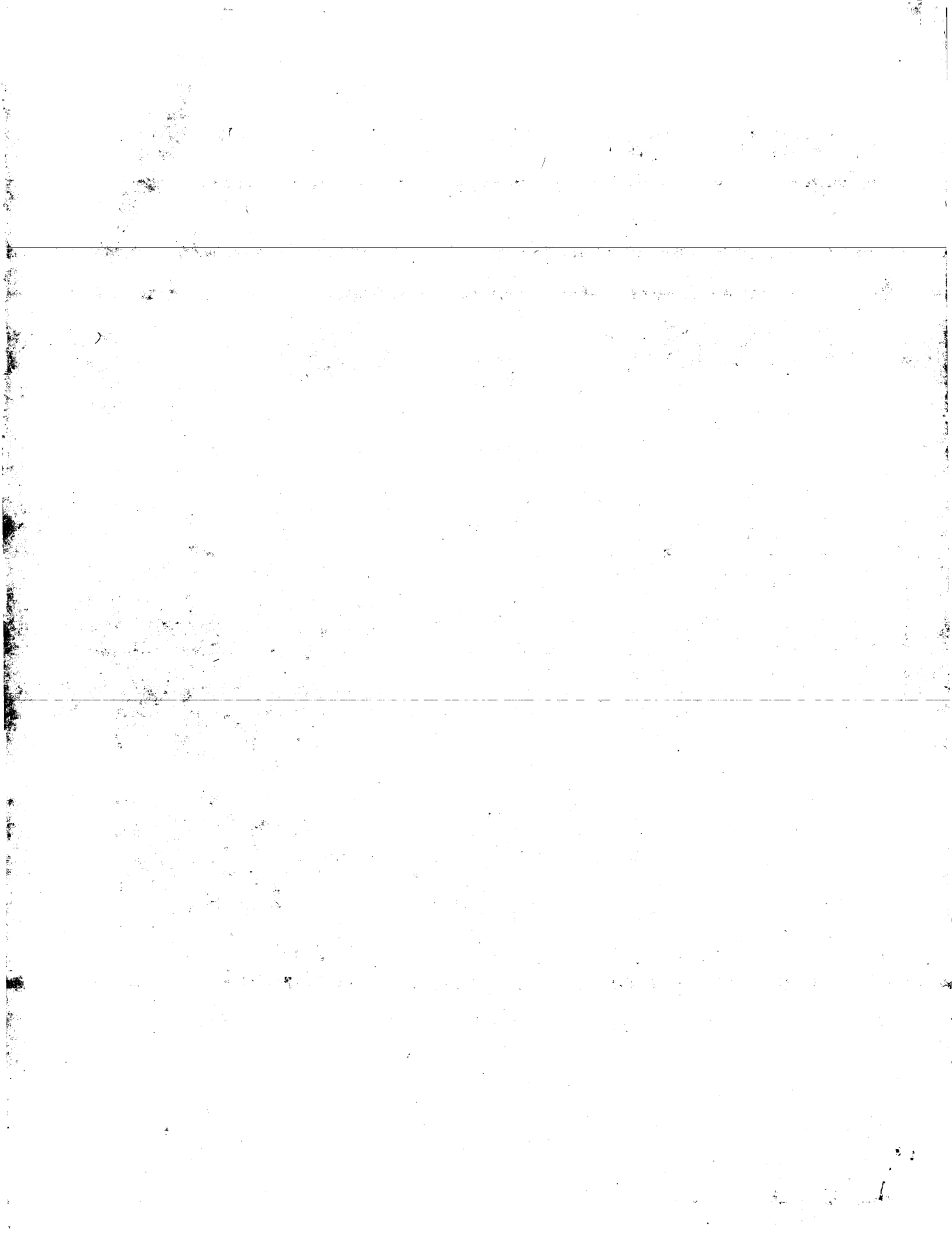
Query Match 69.4%; Score 34; DB 1; Length 4563;
Best Local Similarity 66.7%; Pred. No. 2.8e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVKMDAEFR 10

|||:|:|

Db 1483 EVKIDGQFR 1491

Search completed: September 6, 2001, 16:45:47
Job time: 491 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:51:04 ; Search time 72.75 Seconds
(without alignments)
4.709 Million cell updates/sec

Title: US-09-603-713-4
Perfect score: 49
Sequence: 1 SEVKMDAEFR 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	ID	Description
1	49	100.0	57	1	A4_PIG	Q29023 sus scrofa
2	49	100.0	57	1	A4_URSWA	Q29149 ursus marit
3	49	100.0	58	1	A4_CANFA	Q28280 canis famil
4	49	100.0	58	1	A4_RABIT	Q28748 oryctolagus
5	49	100.0	58	1	A4_SHEEP	Q28757 ovis aries
6	49	100.0	59	1	A4_BOVIN	Q28053 bos taurus
7	49	100.0	751	1	A4_SALSC	Q95241 saimiri sci
8	49	100.0	770	1	A4_HUMAN	P03067 homo sapien
9	44	89.8	770	1	A4_MOUSE	P12023 mus musculu
10	44	89.8	770	1	A4_RAT	P08592 rattus norv
11	34	69.4	3562	1	PGCV_CHICK	Q09053 gallus gall
12	34	69.4	4563	1	APB_HUMAN	P04114 homo sapien
13	33	67.3	927	1	CC15_SCHPO	Q09822 schizosacch
14	32	65.3	354	1	BCPA_CHLLI	Q46135 chlorobium
15	32	65.3	365	1	BCPA_CHLTE	Q46393 chlorobium
16	31	63.3	178	1	YJGA_HAEIN	P45076 haemophilus
17	31	63.3	183	1	YJGA_ECOLI	P26650 escherichia
18	31	63.3	198	1	TFN4_MOUSE	P43488 mus musculu
19	31	63.3	279	1	REFA2_SCHPO	Q92373 schizosacch
20	31	63.3	479	1	Y098_MYCPN	P75353 mycoplasma
21	31	63.3	1024	1	Y075_MYCCE	P73321 mycoplasma
22	30	61.2	78	1	RL31_RICPR	Q9ze47 rickettsia
23	30	61.2	197	1	Y916_HAEIN	Q57483 haemophilus
24	30	61.2	356	1	REF1_BACSU	P45872 bacillus su
25	30	61.2	394	1	EFTU_BUCAI	O31297 buchnera ap
26	30	61.2	419	1	P47K_PSECL	P31521 pseudomonas
27	30	61.2	463	1	YD14_SCHPO	Q92342 schizosacch
28	30	61.2	464	1	SPN5_SCHPO	P48010 schizosacch
29	30	61.2	666	1	ZP2_RABIT	P48829 oryctolagus
30	30	61.2	745	1	ZP2_HUMAN	Q05996 homo sapien
31	30	61.2	745	1	ZP2_MACRA	O77726 macaca radi
32	30	61.2	766	1	METE_YEAST	P05694 saccharomyc
33	30	61.2	827	1	Y4LL_RHISN	P55552 rhizobium s

34	30	61.2	926	1	UVRA_AQUAE	O66911 aquifex aeo
35	30	61.2	1391	1	YLD5_CAEEL	Q03570 caenorhabdi
36	30	61.2	4639	1	DYHC_DROME	P37276 drosophila
37	29	59.2	185	1	RRF_ECOLI	P16174 escherichia
38	29	59.2	221	1	GSPJ_VIBCH	P45776 vibrio chol
39	29	59.2	274	1	YA99_SCHPO	O09787 schizosacch
40	29	59.2	326	1	CC14_CAEEL	P18834 caenorhabdi
41	29	59.2	344	1	SERB_ARCFU	O28142 archaeoglob
42	29	59.2	424	1	EFIA_THEAC	P19486 thermoplasma
43	29	59.2	452	1	F26_YEAST	P32604 saccharomyc
44	29	59.2	478	1	G6PD_BORBU	O51581 borrelia bu
45	29	59.2	491	1	RNG_HAEIN	P45175 haemophilus

ALIGNMENTS

RESULT 1						
A4_PIG						
ID	A4_PIG	STANDARD;	PRT;	57 AA.		
AC	Q29023;					
DT	01-NOV-1997 (Rel. 35, Created)					
DT	01-NOV-1997 (Rel. 35, Last sequence update)					
DT	30-MAY-2000 (Rel. 39, Last annotation update)					
DE	ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN HOMOLOG [CONTAINS: BETA-AMYLOID PROTEIN (BETA-APP) (A-BETA)] (FRAGMENT).					
GN	APP.					
OS	Sus scrofa (Pig).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Cetartiodactyla; Sulina; Suidae; Sus.					
OX	NCBI_TaxID=9823;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	TISSUE=Brain;					
RX	MEDLINE=92017079; PubMed=1656157;					
RA	Johnstone E.M., Chaney M.O., Norris F.H., Pascual R., Little S.P.;					
RT	"Conservation of the sequence of the Alzheimer's disease amyloid peptide in dog, polar bear and five other mammals by cross-species polymerase chain reaction analysis."					
RL	Brain Res. Mol. Brain Res. 10:299-305(1991).					
CC	-!- FUNCTION: FUNCTIONAL NEURONAL RECEPTOR WHICH COUPLES TO INTRACELLULAR SIGNALING PATHWAY THROUGH THE GTP-BINDING PROTEIN G(O) (BY SIMILARITY).					
CC	-!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.					
CC	-!- SIMILARITY: BELONGS TO THE APP FAMILY.					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).					
CC	EMBL; X56127; CAA39592.1; -					
DR	HSSP; P05067; IAML.					
DR	InterPro; IPR001868; -					
DR	PROSITE; PS00319; A4_EXTRA; PARTIAL.					
DR	PROSITE; PS00320; A4_INTRA; PARTIAL.					
KW	Glycoprotein; Amyloid; Neurone; Transmembrane.					
FT	NON_TER 1 1					
FT	CHAIN 6 48					
FT	DOMAIN <1 33					
FT	TRANSMEM 34 57					
FT	NON_TER 57 57					
SEQUENCE	57 AA; 6172 MW; 84209D88EBA82DFA CRC64;					

Query Match 100.0%; Score 49; DB 1; Length 57;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVKMDAEFR 10

```

db      1 SEVKMDAEFR 10
|||||
RESULT 2
A4_URSMA STANDARD; PRT; 57 AA.
Q29149;
01-NOV-1997 (Rel. 35, Created)
01-NOV-1997 (Rel. 35, Last sequence update)
30-MAY-2000 (Rel. 39, Last annotation update)
ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN HOMOLOG [CONTAINS: BETA-AMYLOID
PROTEIN (BETA-APP) (A-BETA)] (FRAGMENT).
APP.
Ursus maritimus (Polar Bear) (Thalarcos maritimus).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Ursidae; Ursus.
NCBI_TaxID=29073;
[1]
SEQUENCE FROM N.A.
TISSUE=Brain;
MEDLINE=92017079; PubMed=1656157;
Johnstone E.M., Chaney M.O., Norris F.H., Pascual R., Little S.P.;
"Conservation of the sequence of the Alzheimer's disease amyloid
peptide in dog, polar bear and five other mammals by cross-species
polymerase chain reaction analysis".
Brain Res. Mol. Brain Res. 10:299-305(1991).
-1- FUNCTION: FUNCTIONAL NEURONAL RECEPTOR WHICH COUPLES TO
INTRACELLULAR SIGNALING PATHWAY THROUGH THE GTP-BINDING PROTEIN
G(i) (BY SIMILARITY).
-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
-1- SIMILARITY: BELONGS TO THE APP FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; X56128; CAA39593.1; -.
HSP; P05067; 1AML.
InterPro; IPR001868; -.
PROSITE; PS00319; A4_EXTRA; PARTIAL.
PROSITE; PS00320; A4_INTRA; PARTIAL.
Glycoprotein; Amyloid; Neurone; Transmembrane.
KW NON_TER 1 1
CHAIN 6 48 BETA-AMYLOID PROTEIN (POTENTIAL).
DOMAIN <1 33 EXTRACELLULAR (POTENTIAL).
TRANSMEM 34 57 POTENTIAL.
FT NON_TER 57
SQ SEQUENCE 57 AA; 6172 MW; 84209D88EA82DFA CRC64;

Query Match 100.0%; Score 49; DB 1; Length 57;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVKMDAEFR 10
|||||
Db 1 SEVKMDAEFR 10
|||||

RESULT 3
A4_CANFA STANDARD; PRT; 58 AA.
ID A4_CANFA
AC Q28280;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN HOMOLOG [CONTAINS: BETA-AMYLOID
DE DE ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN HOMOLOG (FRAGMENT)].

```

```

modified and this statement is not removed. Usage by and for commercial/
entitles requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; X56130; CAA39595.1; -.
HSP; P05067; LAML.
InterPro; IPR001868; -.
PROSITE; PS00319; A4_EXTRA; PARTIAL.
PROSITE; PS00320; A4_INTRA; PARTIAL.
Glycoprotein; Amyloid; Neurone; Transmembrane.
NON_TER 1
CHAIN 6 48 BETA-AMYLOID PROTEIN (POTENTIAL).
DOMAIN <1 33 EXTRACELLULAR (POTENTIAL).
TRANSMEM 34 57 POTENTIAL.
DOMAIN 58 >58 CYTOPLASMIC (POTENTIAL).
NON_TER 58 58
SEQUENCE 58 AA; 6300 MW; F434209D88EBA82D CRC64;
Query Match 100.0%; Score 49; DB 1; Length 58;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SEVKMDAEFR 10
|||||
DDB 1 SEVKMDAEFR 10
|||||
RESULT 6
A4_BOVIN STANDARD; PRT; 59 AA.
IID A4_BOVIN Q28053;
AC Q28053;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN HOMOLOG [CONTAINS: BETA-AMYLOID
DE PROTEIN (BETA-APP) (A-BETA)] (FRAGMENT).
DE APP.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
NCBI_TaxID=9913;
RN [1]
RS SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=92017079; PubMed=1656157;
RT Johnstone E.M., Chaney M.O., Norris F.H., Pascual R., Little S.P.;
RT "Conservation of the sequence of the Alzheimer's disease amyloid
RT peptide in dog, polar bear and five other mammals by cross-species
RT polymerase chain reaction analysis.";
RRL Brain Res. Mol. Brain Res. 10:399-305(1991).
CC -!- FUNCTION: FUNCTIONAL NEURONAL RECEPTOR WHICH COUPLES TO
CC INTRACELLULAR SIGNALING PATHWAY THROUGH THE GTP-BINDING PROTEIN
CC G(O) (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE APP FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entitles requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; X56124; CAA39589.1; -.
EMBL; X56126; CAA39591.1; -.
HSP; P05067; LAML.
InterPro; IPR001868; -.
PROSITE; PS00319; A4_EXTRA; PARTIAL.
PROSITE; PS00320; A4_INTRA; PARTIAL.
Glycoprotein; Amyloid; Neurone; Transmembrane.
NON_TER 1
CHAIN 6 48 BETA-AMYLOID PROTEIN (POTENTIAL).
DOMAIN <1 33 EXTRACELLULAR (POTENTIAL).
TRANSMEM 34 57 POTENTIAL.
DOMAIN 58 >58 CYTOPLASMIC (POTENTIAL).
NON_TER 58 58
SEQUENCE 58 AA; 6300 MW; F434209D88EBA82D CRC64;
Query Match 100.0%; Score 49; DB 1; Length 58;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SEVKMDAEFR 10
|||||
DDB 1 SEVKMDAEFR 10
|||||

```

FT NON_TER 1 1
FT CHAIN 7 49 BETA-AMYLOID PROTEIN (POTENTIAL).
FT DOMAIN <1 34 EXTRACELLULAR (POTENTIAL).
FT TRANSEM 35 58 POTENTIAL.
FT DOMAIN 59 >59 CYTOPLASMIC (POTENTIAL).
FT NON_TER 59 59
SQ SEQUENCE 59 AA: 6414 MW: F43469D488A2E12D CRC64;

Query Match 100.0%; Score 49; DB 1; Length 59;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SEVKMDAEFR 10
Db 2 SEVKMDAEFR 11

RESULT 7
A4_SAISC STANDARD; PRT; 751 AA.
ID A4_SAISC
AC Q95241;
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN PRECURSOR [CONTAINS: BETA-AMYLLOID PROTEIN (BETA-APP) (A-BETA)].
GN APP.
OS Saimiri sciureus (Common squirrel monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae; Saimiri.
OX NCBI_TaxID=9521;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Liver, and Kidney;
RX MEDLINE=96108492; PubMed=8532114;
RA Levy E., Amorim A., Frangione B., Walker L.C.;
RT "Beta-amyloid precursor protein gene in squirrel monkeys with cerebral amyloid angiopathy.";
RL Neurobiol Aging 16:805-808(1995).
CC -!- FUNCTION: FUNCTIONAL NEURONAL RECEPTOR WHICH COUPLES TO INTRACELLULAR SIGNALING PATHWAY THROUGH THE GTP-BINDING PROTEIN G(O).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- DOMAIN: THE CLATHRIN-BINDING SITE IS ESSENTIAL FOR ITS ASSOCIATION WITH X11-ALPHA, -BETA, AND -GAMMA. THE SEQUENCE SPECIFIC RECOGNITION EXTENDS TO PEPTIDE RESIDUES THAT ARE C-TERMINAL TO THE NPXY MOTIF. THIS INTERACTION APPEARS TO BE INDEPENDENT OF PHOSPHORYLATION (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE APP FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL; S81024; AAD14347.1;
CC InterPro; IPR001255;
CC InterPro; IPR001868;
CC InterPro; IPR002223;
CC Pfam; PF00014; Kunitz_BPTI; 1.
CC PRINTS; PR00203; AMYLOIDA4.
CC PRINTS; PR00204; BETAAMYLOID.
CC PRINTS; PR00759; BASICTPASE.
CC PROSITE; PS00319; A4_EXTRA; 1.
CC PROSITE; PS00320; A4_INTRA; 1.
CC PROSITE; PS00280; BPTI_KUNITZ_1; 1.
CC PROSITE; PS00279; BPTI_KUNITZ_2; 1.
CC Glycoprotein; Amyloid; Neurone; Transmembrane; Alternative splicing; Signal; Serine protease inhibitor.

FT SIGNAL 1 17
FT CHAIN 18 751
FT CHAIN 653 695 BETA-AMYLOID PROTEIN (POTENTIAL).
FT DOMAIN 18 680 EXTRACELLULAR (POTENTIAL).
FT TRANSEM 681 704 POTENTIAL.
FT DOMAIN 705 751 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 287 345 BPTI/KUNITZ INHIBITOR.
FT SITE 740 743 CLATHRIN-BINDING (BY SIMILARITY).
FT ACT_SITE 301 302 REACTIVE BOND.
FT DISULFID 291 341
FT DISULFID 300 324 BY SIMILARITY.
FT DISULFID 316 337 BY SIMILARITY.
FT CARBOHYD 523 523 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 552 552 N-LINKED (GLCNAC. . .) (PROBABLE).
SQ SEQUENCE 751 AA: 84893 MW: 6C3E431089569049 CRC64;

Query Match 100.0%; Score 49; DB 1; Length 751;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SEVKMDAEFR 10
Db 648 SEVKMDAEFR 657

RESULT 8
A4_HUMAN STANDARD; PRT; 770 AA.
ID A4_HUMAN
AC P03067; P09000; Q16011;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN PRECURSOR (PROTEASE NEXIN-II) (PN-II) (APPI) [CONTAINS: BETA-AMYLOID PROTEIN (BETA-APP) (A-BETA)].
GN APP OR A4 OR CVAP OR ADL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Brain;
RX MEDLINE=87144572; PubMed=2881207;
RA Kang J., Lemaire H.-G., Unterbeck A., Salbaum J.M., Masters C.L., Grzeschik K.-H., Multhaup G., Beyreuther K., Mueller-Hill B.;
RT "The precursor of Alzheimer's disease amyloid A4 protein resembles a cell-surface receptor.";
RL Nature 325:733-736(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88122639; PubMed=2893289;
RA Ponte P., Gonzalez-Dewhitt P., Schilling J., Miller J., Hsu D., Greenberg B., Davis K., Wallace W., Lieberburg I., Fuller F., Cordell B.;
RT "A new A4 amyloid mRNA contains a domain homologous to serine proteinase inhibitors.";
RL Nature 331:525-527(1988).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=89128427; PubMed=2783775;
RA Lemaire H.-G., Salbaum J.M., Multhaup G., Kang J., Bayney R.M., Unterbeck A., Beyreuther K., Mueller-Hill B.;
RT "The PreA4(695) precursor protein of Alzheimer's disease A4 amyloid is encoded by 16 exons.";
RL Nucleic Acids Res. 17:517-522(1989).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=97263807; PubMed=9108164;
RA Hattori M., Tsukahara F., Furuhata Y., Tanahashi H., Hirose M., Saito M., Tsukuni S., Sakaki Y.;
RT "A novel method for making nested deletions and its application for sequencing of a 300 kb region of human APP locus.";

- RL Nucleic Acids Res. 25:1802-1808(1997).
 RN [5]
 RP SEQUENCE OF 286-345 AND 365-366 FROM N.A.
 RX MEDLINE-88122640; PubMed-2893290;
 RA Tanzi R.E., McClatchey A.I., Lamperti E.D., Villa-Komaroff L.,
 Gusella J.F., Neve R.L.;
 RT "Protease inhibitor domain encoded by an amyloid protein precursor
 RT mRNA associated with Alzheimer's disease.";
 RL Nature 331:528-530(1988).
 RN [6]
 RP SEQUENCE OF 287-367 FROM N.A.
 RX MEDLINE-88122641; PubMed-2893291;
 RA Kitaguchi N., Takahashi Y., Tokushima Y., Shiojiri S., Ito H.;
 RT "Novel precursor of Alzheimer's disease amyloid protein shows
 RT protease inhibitory activity.";
 RL Nature 331:530-532(1988).
 RN [7]
 RP SEQUENCE OF 284-289 AND 365-770 FROM N.A.
 RX MEDLINE-87231971; PubMed-3035574;
 RA Robakis N.K., Ramakrishna N., Wolfe G., Wisniewski H.M.;
 RT "Molecular cloning and characterization of a cDNA encoding the
 RT cerebrovascular and the neuritic plaque amyloid peptides.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:4190-4194(1987).
 RN [8]
 RP SEQUENCE OF 507-770 FROM N.A.
 RX MEDLINE-88124954; PubMed-2893379;
 RA Zain S.B., Salim M., Chou W.G., Sajdel-Sulkowska E.M., Majocha R.E.,
 RA Marotta C.A.;
 RT "Molecular cloning of amyloid cDNA derived from mRNA of the Alzheimer
 RT disease brain: coding and noncoding regions of the fetal precursor
 RT mRNA are expressed in the cortex.";
 RL Proc. Natl. Acad. Sci. U.S.A. 85:929-933(1988).
 RN [9]
 RP SEQUENCE OF 672-681.
 RX MEDLINE-88035004; PubMed-3312495;
 RA Partridge W.M., Vinters H.V., Yang J., Eisenberg J., Choi T.B.,
 RA Tourtellotte W.W., Huebner V., Shively J.E.;
 RT "Amyloid angiopathy of Alzheimer's disease: amino acid composition
 RT and partial sequence of a 4,200-dalton peptide isolated from cortical
 RT microvessels.";
 RL J. Neurochem. 49:1394-1401(1987).
 RN [10]
 RP SEQUENCE OF 739-770 FROM N.A.
 RX MEDLINE-90236318; PubMed-2110105;
 RA Yoshikai S.-I., Sasaki H., Doh-Ura K., Furuya H., Sakaki Y.;
 RT "Genomic organization of the human amyloid beta-protein precursor
 RT gene.";
 RL Gene 87:257-263(1990).
 RN [11]
 RP SEQUENCE OF 1-10 FROM N.A.
 RX MEDLINE-89016647; PubMed-3140222;
 RA Schon E.A., Mita S., Sadlock J., Herbert J.;
 RT "A cDNA specifying the human amyloid beta precursor protein (ABPP)
 RT encodes a 95-kDa polypeptide.";
 RL Nucleic Acids Res. 16:9351-9351(1988).
 RN [12]
 RP SEQUENCE OF 18-50.
 RX MEDLINE-87250462; PubMed-3597385;
 RA van Nostrand W.E., Cunningham D.D.;
 RT "Purification of protease nexin II from human fibroblasts.";
 RL J. Biol. Chem. 262:8508-8514(1987).
 RN [13]
 RP IDENTITY OF APP WITH NEXIN-II.
 RX MEDLINE-89384866; PubMed-2506449;
 RA Oltsdorf T., Fritz L.C., Schenk D.B., Lieberburg I.,
 RA Johnson-Wood K.L., Beattie E.C., Ward P.J., Blacher R.W., Dovey H.F.,
 RA Sinha S.;
 RT "The secreted form of the Alzheimer's amyloid precursor protein with
 RT the Kunitz domain is protease nexin-II.";
 RL Nature 341:144-147(1989).
 RN [14]
 RP PROTEASE-SPECIFICITY OF INHIBITOR DOMAIN.
 RX MEDLINE-90211252; PubMed-1969731;
 RA Kido H., Fukutomi A., Schilling J., Wang Y., Cordell B., Katunuma N.;
 RT "Protease-specificity of Kunitz inhibitor domain of Alzheimer's
 RT disease amyloid protein precursor.";
 RL Biochem. Biophys. Res. Commun. 167:716-721(1990).
 RN [15]
 RP COMPLEX WITH G(O).
 RX MEDLINE-93188965; PubMed-8446172;
 RA Nishimoto T., Okamoto T., Matsuura Y., Takahashi S., Okamoto T.,
 RA Murayama Y., Ogata E.;
 RT "Alzheimer amyloid protein precursor complexes with brain GTP-binding
 RT protein G(o).";
 RL Nature 362:75-79(1993).
 RN [16]
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 28-133.
 RX MEDLINE-95215582; PubMed-10201399;
 RA Rossjohn J., Cappai R., Feil S.C., Henry A., McKinstry W.J.,
 RA Galatis D., Hesse L., Multhaup G., Beyreuther K., Masters C.L.,
 RA Parker M.W.;
 RT "Crystal structure of the N-terminal, growth factor-like domain of
 RT Alzheimer amyloid precursor protein.";
 RL Nat. Struct. Biol. 6:327-331(1999).
 RN [17]
 RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF 287-344.
 RX MEDLINE-91104913; PubMed-2125487;
 RA Hynes T.R., Randal M., Kennedy L.A., Eigenbrot C., Kosiakof A.A.;
 RT "X-ray crystal structure of the protease inhibitor domain of
 RT Alzheimer's amyloid beta-protein precursor.";
 RL Biochemistry 29:10018-10022(1990).
 RN [18]
 RP STRUCTURE BY NMR OF 289-344.
 RX MEDLINE-92031488; PubMed-1718421;
 RA Heald S.L., Tilton R.F. Jr., Hammond L.S., Lee A., Bayney R.M.,
 RA Kamark M.E., Ramabhadran T.V., Dreyer R.N., Davis G., Unterbeck A.,
 RA Tamburini F.P.;
 RT "Sequential NMR resonance assignment and structure determination of
 RT the Kunitz-type inhibitor domain of the Alzheimer's beta-amyloid
 RT precursor protein.";
 RL Biochemistry 30:10467-10478(1991).
 RN [19]
 RP STRUCTURE BY NMR OF 672-699.
 RX MEDLINE-94281210; PubMed-7516706;
 RA Talaous J., Marciniowski K.J., Klopman G., Zagorski M.G.;
 RT "Solution structure of residues 1-28 of the amyloid beta-peptide.";
 RL Biochemistry 33:7788-7796(1994).
 RN [20]
 RP STRUCTURE BY NMR OF 696-706.
 RX MEDLINE-97128622; PubMed-8973180;
 RA Kohno T., Kobayashi K., Maeda T., Sato K., Takashima A.;
 RT "Three-dimensional structures of the amyloid beta peptide (25-35) in
 RT membrane-mimicking environment.";
 RL Biochemistry 35:16094-16104(1996).
 RN [21]
 RP STRUCTURE BY NMR OF 672-711.
 RX MEDLINE-98359783; PubMed-9693002;
 RA Coles M., Bicknell W., Watson A.A., Craik D.J.;
 RT "Solution structure of amyloid beta-peptide(1-40) in a water-micelle
 RT environment. Is the membrane-spanning domain where we think it is?";
 RL Biochemistry 37:11064-11077(1998).
 RN [22]
 RP STRUCTURE BY NMR OF 672-699.
 RX MEDLINE-20400066; PubMed-10940222;
 RA Poulsen S.-A., Watson A.A., Craik D.J.;
 RT "Solution structures in aqueous SDS micelles of two amyloid beta
 RT peptides of Abeta(1-28) mutated at the alpha-secretase cleavage
 RT site.";
 RL J. Struct. Biol. 130:142-152(2000).
 RN [23]
 RP SIGNAL SEQUENCE CLEAVAGE SITE, AND TOPOLOGY.
 RX MEDLINE-88296437; PubMed-2900137;
 RA Dyrks T., Weidemann A., Multhaup G., Salbaum J.M., Lemaire H.-G.,
 RA Kang J., Mueller-Hill B., Masters C.L., Beyreuther K.;
 RT "Identification, transmembrane orientation and biogenesis of the

```

RT amyloid A4 precursor of Alzheimer's disease."
RL EMBL J. 7:949-957(1988).
RN [24]
RP REVIEW.
RX MEDLINE=92271194; PubMed=1589757;
RA Kosik K.S.;
RT "Alzheimer's disease: a cell biological perspective.";
RL Science 256:780-783(1992).
RN [25]

Query Match 100.0%; Score 49; DB 1; Length 770;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SEVKMDAEFR 10
Db 667 SEVKMDAEFR 676

RESULT 9
A4_MOUSE
ID A4_MOUSE STANDARD; PRT; 770 AA.
AC P12023;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN HOMOLOG PRECURSOR
DE (AMYLOIDOGENIC GLYCOPROTEIN) (AG).
DS APP.
GN Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE OF 1-289 AND 365-770 FROM N.A.
RX STRAIN=BALB/C; TISSUE=Brain;
RC MEDLINE=92096458; PubMed=1756177;
RA De Strooper B., van Leuven F., van den Bergh H.;
RT "The amyloid beta protein precursor or proteinase nexin II from mouse
RT is closer related to its human homolog than previously reported.";
RL Biochim. Biophys. Acta 1129:141-143(1991).
RN [2]
RP SEQUENCE OF 1-289 AND 365-770 FROM N.A.
RX TISSUE=Brain;
RC MEDLINE=88106489; PubMed=3322280;
RA Yamada T., Sasaki H., Furuya H., Miyata T., Goto I., Sakaki Y.;
RT "Complementary DNA for the mouse homolog of the human amyloid beta
RT protein precursor.";
RL Biochem. Biophys. Res. Commun. 149:665-671(1987).
RN [3]
RP REVISIONS.
RA Yamada T.;
RL Submitted (MAR-1988) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 289-364 FROM N.A.
RX STRAIN=CD-1; TISSUE=Placenta;
RC MEDLINE=89345111; PubMed=2569710;
RA Fukuchi K., Martin G.M., Deeb S.S.;
RT "Sequence of the protease inhibitor domain of the A4 amyloid protein
RT precursor of Mus domesticus.";
RL Nucleic Acids Res. 17:5396-5396(1989).
RN [5]
RP SEQUENCE OF 1-19 FROM N.A.
RX MEDLINE=92209998; PubMed=1555768;
RA Izumi R., Yamada T., Yoshikai S.I., Sasaki H., Hattori M.,
RA Sakai Y.;
RT "Positive and negative regulatory elements for the expression of the
RT Alzheimer's disease amyloid precursor-encoding gene in mouse.";
RL Gene 112:189-195(1992).
RN [6]
RP SEQUENCE OF 281-380 FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=Brain, and Kidney;
RX MEDLINE=89149813; PubMed=2493250;

```

```

RA Yamada T., Sasaki H., Dohura K., Goto I., Sakaki Y.;
RT "Structure and expression of the alternatively-spliced forms of mRNA
RT for the mouse homolog of Alzheimer's disease amyloid beta protein
RT precursor.";
RL Biochem. Biophys. Res. Commun. 158:906-912(1989).
RN [7]
RP SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- APP(563), APP(695), APP(751) AND APP(770) (SHOWN HERE); WHICH ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: AAA(770) IS EXPRESSED IN KIDNEY. AAA(751) IS
CC WIDELY EXPRESSED. AAA(695) IS EXPRESSED IN BRAIN, KIDNEY AND
CC LIVER.
CC -!- DOMAIN: THE CLATHRIN-BINDING SITE IS ESSENTIAL FOR ITS ASSOCIATION
CC WITH X11-ALPHA, -BETA, AND -GAMMA. THE SEQUENCE SPECIFIC
CC RECOGNITION EXTENDS TO PEPTIDE RESIDUES THAT ARE C-TERMINAL TO THE
CC NPXY MOTIF. THIS INTERACTION APPEARS TO BE INDEPENDENT OF
CC PHOSPHORYLATION (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE APP FAMILY.
CC -!- SIMILARITY: APP(751) AND APP(770) CONTAIN A PROTEASE INHIBITOR
CC DOMAIN BELONGING TO THE BPTI/KUNITZ FAMILY OF INHIBITORS.
RN [8]
RP This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
RN [9]
RP EMBL; X59379; ; NOT_ANNOTATED_CDS.
DR EMBL; M18373; AAA37139.1; -
DR EMBL; X15210; CAA33280.1; -
DR EMBL; D10603; BAA01456.1; -
DR EMBL; M24397; AAA39929.1; -
DR PIR; A27485; A27485
DR PIR; S04855; S04855
DR PIR; S19727; S19727
DR MGD; MGI:88059; App.
DR InterPro; IPR001255; -
DR InterPro; IPR001868; -
DR InterPro; IPR002223; -
DR Pfam; PF00014; Kunitz_BPTI; 1.
DR PRINTS; PR00203; AMYLOIDA4.
DR PRINTS; PR00204; BETAAMYLOID.
DR PRINTS; PR00759; BASICPTASE.
DR PROSITE; PS00319; A4_EXTRA; 1.
DR PROSITE; PS00320; A4_INTRA; 1.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 1.
DR PROSITE; PS00279; BPTI_KUNITZ_2; 1.
DR KW - Glycoprotein; Amyloid; Neurone; Transmembrane; Signal;
KW Alternative splicing; Serine protease inhibitor.
FT SIGNAL 1 17
FT CHAIN 18 770
FT DOMAIN 18 699
FT TRANSMEM 700 723
FT DOMAIN 724 770
FT DOMAIN 673 715
FT DOMAIN 287 345
FT SITE 759 762
FT DISULFID 291 341
FT DISULFID 300 324
FT DISULFID 316 337
FT CARBOHYD 542 542
FT CARBOHYD 571 571
FT VARSPLIC 289 289
FT VARSPLIC 290 364
FT VARSPLIC 346 380
FT SEQUENCE 770 AA; 86752 MW; 26C50DE0890CAFA7A CRC64;
SQ

```

Query Match 89.8%; Score 44; DB 1; Length 770;
Best Local Similarity 100.0%; Pred. No. 0.22;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVKMDAEF 9
 |||||
 DB 667 SEVKMDAEF 675

RESULT 10
 ID A4_RAT STANDARD; PRT: 770 AA.
 AC P08592;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN HOMOLOG PRECURSOR
 DE (AMYLOIDGENIC GLYCOPROTEIN) (AG).
 GN APP.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE OF 1-289 AND 365-770 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=88312583; PubMed=2900758;
 RA Shivers B.D., Hilbich C., Multhaup G., Salbaum J.M., Beyreuther K.,
 RA Seeburg P.H.;
 RT "Alzheimer's disease amyloidogenic glycoprotein: expression pattern
 RT in rat brain suggests a role in cell contact.";
 RL EMBO J. 7:1365-1370(1988).
 RN [2]
 RP SEQUENCE OF 289-364 FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=89183625; PubMed=2648331;
 RA Kang J., Mueller-Hill B.;
 RT "The sequence of the two extra exons in rat preA4.";
 RL Nucleic Acids Res. 17:2130-2130(1989).
 CC -|- SUBCELLULAR LOCATION: SIX FORMS OF APP ARE FOUND; APP(395),
 CC APP(563), APP(695), APP(751) AND APP(770) (SHOWN HERE); WHICH ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -|- DOMAIN: THE CLATHRIN-BINDING SITE IS ESSENTIAL FOR ITS ASSOCIATION
 CC WITH X11-ALPHA, -BETA, AND -GAMMA. THE SEQUENCE SPECIFIC
 CC RECOGNITION EXTENDS TO PEPTIDE RESIDUES THAT ARE C-TERMINAL TO THE
 CC NPXY MOTIF. THIS INTERACTION APPEARS TO BE INDEPENDENT OF
 CC PHOSPHORYLATION (BY SIMILARITY).
 CC -|- SIMILARITY: BELONGS TO THE APP FAMILY.
 CC -|- SIMILARITY: APP(751) AND APP(770) CONTAIN A PROTEASE INHIBITOR
 CC DOMAIN BELONGING TO THE BPTI/KUNITZ FAMILY OF INHIBITORS.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X07648; CAA30488.1; -;
 DR EMBL: X14056; CAA32229.1; -;
 DR PIR: S00550; S00550.
 DR PIR: S03607; S03607.
 DR InterPro: IPR001255; -;
 DR InterPro: IPR001868; -;
 DR InterPro: IPR002223; -;
 DR Pfam: PF00014; Kunitz_BPTI; 1.
 DR PRINTS: PR00203; AMYLOIDA4.
 DR PRINTS: PR00204; BETAAMYLOID.
 DR PRINTS: PR00759; BASICPTASE.
 DR PROSITE: PS00319; A4_EXTRA; 1.
 DR PROSITE: PS00320; A4_INTRA; 1.
 DR PROSITE: PS00280; BPTI_KUNITZ_1; 1.
 DR PROSITE: PS00279; BPTI_KUNITZ_2; 1.

KW Glycoprotein; Amyloid; Neurone; Transmembrane; Signal;
 KW Alternative splicing; Serine protease inhibitor.
 FT SIGNAL 1 17
 FT CHAIN 18 770
 FT ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN
 FT HOMOLOG.
 FT DOMAIN 18 699
 FT TRANSMEM 700 723
 FT DOMAIN 724 770
 FT DOMAIN 673 715
 FT DOMAIN 287 345
 FT SITE 759 762
 FT DISULFID 291 341
 FT DISULFID 300 324
 FT DISULFID 316 337
 FT CARBOHYD 542 542
 FT CARBOHYD 571 571
 FT VARSPLIC 289 289
 FT VARSPLIC 290 364
 FT SEQUENCE 770 AA; 86704 MW; C26C9D6B2D929A7 CRC64;
 SQ

Query Match 89.8%; Score 44; DB 1; Length 770;
 Best Local Similarity 100.0%; Pred. No. 0.22;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVKMDAEF 9
 |||||
 DB 667 SEVKMDAEF 675

RESULT 11

PGCV_CHICK
 ID PGCV_CHICK STANDARD; PRT: 3562 AA.
 AC Q90953; Q90945;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE VERSICAN CORE PROTEIN PRECURSOR (LARGE FIBROBLAST PROTEOGLYCAN)
 DE (CHONDROITIN SULFATE PROTEOGLYCAN CORE PROTEIN 2) (PG-M).
 GN CPSC2.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=WHITE LEGHORN; TISSUE=Limb bud;
 RX MEDLINE=93300846; PubMed=8314802;
 RA Shinomura T., Nishida Y., Ito K., Kimata K.;
 RT "cDNA cloning of PG-M, a large chondroitin sulfate proteoglycan
 RT expressed during chondrogenesis in chick limb buds. Alternative
 RT spliced multiforms of PG-M and their relationships to versican.";
 RL J. Biol. Chem. 268:14461-14469(1993).
 CC -|- FUNCTION: MAY PLAY A ROLE IN INTERCELLULAR SIGNALING AND IN
 CC CONNECTING CELLS WITH THE EXTRACELLULAR MATRIX. MAY TAKE PART IN
 CC THE REGULATION OF CELL MOTILITY, GROWTH AND DIFFERENTIATION. BINDS
 CC HYALURONIC ACID.
 CC -|- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX.
 CC -|- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; V0 (SHOWN HERE) AND
 CC V1; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -|- TISSUE SPECIFICITY: PRECHONDROGENIC CONDENSATION AREA OF
 CC DEVELOPING LIMB BUDS.
 CC -|- DEVELOPMENTAL STAGE: DISAPPEARS AFTER THE CARTILAGE DEVELOPMENT
 CC (BY SIMILARITY).
 CC -|- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
 CC -|- SIMILARITY: CONTAINS 2 LINK DOMAINS.
 CC -|- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
 CC -|- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC -|- SIMILARITY: CONTAINS 1 SUSHI (SCR) REPEAT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

RL EMBO J. 5:3495-3507(1986).
 RN [6]
 RP SEQUENCE OF 709-906 FROM N.A.
 RX MEDLINE=85270450; PubMed=3860836;
 RA Deeb S.S., Motulsky A.G., Albers J.J.;
 RT "A partial cDNA clone for human apolipoprotein B.";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:4983-4986(1985).
 RN [7]
 RP SEQUENCE OF 3056-3159 FROM N.A.
 RX MEDLINE=86041888; PubMed=3903660;
 RA Mehrabian M., Schumaker V.N., Fareed G.C., West R., Johnson D.F.,
 RA Kirchgessner T.G., Lin H.-C., Wang X., Ma Y., Mendiaz E., Lusis A.J.;
 RT "Human apolipoprotein B: identification of cDNA clones and
 RT characterization of mRNA.";
 RL Nucleic Acids Res. 13:6937-6953(1985).
 RN [8]
 RP SEQUENCE OF 1937-2018 AND 3811-4334 FROM N.A.
 RX MEDLINE=86093680; PubMed=3841204;
 RA Carlsson P., Olofsson S.O., Bondjers G., Darnfors C., Wiklund O.,
 RA Bjursell G.;
 RT "Molecular cloning of human apolipoprotein B cDNA.";
 RL Nucleic Acids Res. 13:8813-8826(1985).
 RN [9]
 RP SEQUENCE OF 3109-4563 FROM N.A.
 RX MEDLINE=85300528; PubMed=2994225;
 RA Knott T.J., Rall S.C. Jr., Innerarity T.L., Jacobson S.F.,
 RA Urdea M.S., Levy-Wilson B., Powell L.M., Pease R.J., Eddy R.,
 RA Nakai H., Byers M., Priestley L.M., Robertson E., Rall L.B.,
 RA Betsholtz C., Shows T.B., Mahley R.W., Scott J.;
 RT "Human apolipoprotein B: structure of carboxyl-terminal domains,
 RT sites of gene expression, and chromosomal localization.";
 RL Science 230:37-43(1985).
 RN [10]
 RP SEQUENCE OF 1-291 FROM N.A.
 RX MEDLINE=86149325; PubMed=3513177;
 RA Protter A.A., Hardman D.A., Schilling J.W., Miller J., Appleby V.,
 RA Chen G.C., Kirsher S.W., McEnroe G., Kane J.P.;
 RT "Isolation of a cDNA clone encoding the amino-terminal region of
 RT human apolipoprotein B.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:1467-1471(1986).
 RN [11]
 RP SEQUENCE OF 1-1670 FROM N.A.
 RX MEDLINE=86287319; PubMed=3461454;
 RA Protter A.A., Hardman D.A., Sato K.Y., Schilling J.W.,
 RA Yamana M., Hort Y.J., Hjerrild K.A., Chen G.C., Kane J.P.;
 RT "Analysis of cDNA clones encoding the entire B-26 region of human
 RT apolipoprotein B.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:5678-5682(1986).
 RN [12]
 RP PARTIAL SEQUENCE, AND IDENTIFICATION (APO-B48).
 RX MEDLINE=88018019; PubMed=3659919;
 RA Chen S.-H., Habib G., Yang C.-H., Gu Z.-W., Lee B.R., Weng S.-H.,
 RA Silberman S.R., Cal S.-J., Deslypere J.P., Rosseneu M.,
 RA Gotto A.M. Jr., Li W.-H., Chan L.;
 RT "Apolipoprotein B-48 is the product of a messenger RNA with an organ-
 RT specific in-frame stop codon.";
 RL Science 238:363-366(1987).
 RN [13]
 RP DOMAINS.
 RX MEDLINE=87039351; PubMed=3773987;
 RA Knott T.C., Pease R.J., Powell L.M., Wallis S.C., Rall S.C. Jr.,
 RA Innerarity T.L., Blackhart B., Taylor W.R., Marcel Y., Milne R.,
 RA Johnson D., Fuller M., Lusis A.J., McCarthy B.J., Mahley R.W.,
 RA Levy-Wilson B., Scott J.;
 RT "Complete protein sequence and identification of structural domains
 RT of human apolipoprotein B.";
 RL Nature 323:734-738(1986).
 RN [14]
 RP DOMAINS.
 RA Yang C.-Y., Chen S.-H., Gianturco S.H., Bradley W.A., Sparrow J.T.,
 RA Tanamura M., Li W.-H., Sparrow D.A., Deloof H., Rosseneu M.,
 RA Lee F.-S., Gu Z.-W., Gotto A.M. Jr., Chan L.;
 RT "Sequence, structure, receptor-binding domains and internal repeats
 of human apolipoprotein B-100.";
 RL Nature 323:738-742(1986).
 RN [15]
 RP CALCULUM-BINDING DATA.
 RX MEDLINE=86242245; PubMed=3087360;
 RA Dashti N., Lee D.M., Mok T.;
 RT "Apolipoprotein B is a calcium binding protein.";
 RL Biochem. Biophys. Res. Commun. 137:493-499(1986).
 RN [16]
 RP VARIANT SER-4338.
 RX MEDLINE=91071750; PubMed=1979313;
 RA Navajas M., Laurent A.-M., Moreel J.-F., Ragab A., Cambou J.-P.,
 RA Cuny G., Cambien F., Roizes G.;
 RT "Detection by denaturing gradient gel electrophoresis of a new
 RT polymorphism in the apolipoprotein B gene.";
 RL Hum. Genet. 86:91-93(1990).
 RN [17]
 RP VARIANT FDB GLN-3527.
 RX MEDLINE=89098975; PubMed=2563166;
 RA Soria L.F., Ludwig E.H., Clarke H.R.G., Vega G.L., Grundy S.M.,
 RA McCarthy B.J.;
 RT "Association between a specific apolipoprotein B mutation and
 RT familial defective apolipoprotein B-100.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:587-591(1989).
 RN [18]
 RP VARIANT LEU-2739.
 RX MEDLINE=91016974; PubMed=2216805;
 RA Huang L.-S., Gavish D., Breslow J.L.;
 RT "Sequence polymorphism in the human apoB gene at position 8344.";
 RL Nucleic Acids Res. 18:5922-5922(1990).
 RN [19]
 RP VARIANT FDB CYS-3558.
 RX MEDLINE=95190020; PubMed=7883971;
 RA Pullinger C.R., Hennessy L.K., Chatterton J.E., Liu W., Love J.A.,
 RA Mendel C.M., Frost P.H., Malloy M.J., Schumaker V.N., Kane J.P.;
 RT "Familial ligand-defective apolipoprotein B. Identification of a new
 RT mutation that decreases LDL receptor binding affinity.";
 RL J. Clin. Invest. 95:1225-1234(1995).
 RN [20]
 RP VARIANTS L-1437; S-1914; K-2566; T-3121; A-3945; M-4128 AND T-4481.
 RX MEDLINE=97044521; PubMed=889592;
 RA Poirier O., Ricard S., Behague I., Souriau C., Evans A.E.,
 RA Arveller D., Marques-Vidal P., Luc G., Roizes G., Cambien F.;
 RT "Detection of new variants in the apolipoprotein B (Apo B) gene by
 RT PCR-SSCP.";
 RL Hum. Mutat. 8:282-285(1996).
 RN [21]
 RP VARIANTS FDB GLN-3527 AND CYS-3558.
 RX MEDLINE=97403938; PubMed=9259199;
 RA Rabes J.P., Varret M., Saint-Jore B., Erlich D., Jondeau G.,
 RA Krempf M., Giraudet P., Junien C., Boileau C.;
 RT "Familial ligand-defective apolipoprotein B-100: simultaneous
 RT detection of the ARG3500-->GLN and ARG3531-->CYS mutations in a
 RT French population.";
 RL Hum. Mutat. 10:160-163(1997).
 RN [22]
 RP VARIANTS S-1914; R-1923; L-2739; D-3319; T-3427; Q-3432 AND I-3921.
 RX MEDLINE=98141125; PubMed=9450296;
 RA Leren T.P., Bakken K.S., Hoel V., Hjermann I., Berg K.;
 RT "Screening for mutations of the apolipoprotein B gene causing
 RT hypocholesterolemia.";
 RL Hum. Genet. 102:44-49(1998).
 CC -!- FUNCTION: APOLOPROTEIN B IS A MAJOR PROTEIN CONSTITUENT OF
 CC CHYLOMICRONS, VLDL AND LDL. IT FUNCTIONS AS A RECOGNITION SIGNAL
 CC FOR THE CELLULAR BINDING AND INTERNALIZATION OF LDL PARTICLES BY
 CC THE APOB/E RECEPTOR.
 CC -!- DISEASE: DEFECTS IN APOB ARE A CAUSE OF FAMILIAL LIGAND-DEFECTIVE
 CC APOLOPROTEIN B-100 (FDB). IT IS A DOMINANTLY INHERITED DISORDER
 CC OF LIPOPROTEIN METABOLISM LEADING TO HYPERCHOLESTEROLEMIA AND
 CC INCREASED PRONESS TO CORONARY ARTERY DISEASE (CAD).
 CC THE PLASMA CHOLESTEROL LEVELS ARE DRAMATICALLY ELEVATED DUE TO
 CC IMPAIRED CLEARANCE OF LDL PARTICLES BY DEFECTIVE APOB/E RECEPTORS.
 CC -!- DISEASE: DEFECTS IN APOB ASSOCIATED WITH OTHER GENE EFFECTS

CC (POLYGENIC) CAN BE THE CAUSE OF HYPOCHOLESTEROLEMIA.
CC -1- MISCELLANEOUS: APO B-48 WHICH IS SYNTHESIZED ONLY BY THE
CC INTESTINE, AND FOUND IN CHYLOMICRONS, IS A SHORTENED FORM OF APO

Query Match 69.4%; Score 34; DB 1; Length 4563;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 EVKMDAEFR 10
Db 1483 EVKIDQFR 1491

RESULT 13
CC15_SCHPO STANDARD; PRT; 927 AA.
ID CC15_SCHPO STANDARD; PRT; 927 AA.
AC 009822; 014365;
DT 01-FEB-1996 (Rel. 33, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CELL DIVISION CONTROL PROTEIN 15.
GN CDC15 OR SPAC20G8.05C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
CC Schizosaccharomycetales; Schizosaccharomycetaceae;
CC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=95360987; PubMed=7634333;
RA Fankhauser C., Raymond A., Cerutti L., Utzig S., Hofmann K.,
RA Simanis V.;
RT "The S. pombe cdc15 gene is a key element in the reorganization of F-
RT actin at mitosis.";
RL Cell 82:435-444(1995).
RN [2]
RP REVISTONS TO N-TERMINUS.
RA Fankhauser C., Raymond A., Cerutti L., Utzig S., Hofmann K.,
RA Simanis V.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Badcock K., Churcher C.M., Wood V., Barrell B.G., Rajandream M.A.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: AFTER THE ONSET OF MITOSIS, FORMS A RING-LIKE STRUCTURE
CC WHICH CO-LOCALIZES WITH THE MEDIAL ACTIN RING. APPEARS TO MEDIATE
CC CYTOSKELETAL REARRANGEMENTS REQUIRED FOR CYTOKINESIS. ESSENTIAL
CC FOR VIABILITY.
CC -1- DEVELOPMENTAL STAGE: PEAKS IN EARLY MITOSIS BEFORE SEPTATION.
CC -1- DOMAIN: THE N-TERMINAL REGION IS IN A COILED COIL STRUCTURE.
CC -1- PTM: PHOSPHORYLATED.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: SOME, TO S.POMBE SPBCL11.02 AND SPAC704.02C.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

EMBL; X86179; CAA60115.1; -;
DR EMBL; Z95334; CAB08599.2; -;
DR HSSP; P07751; ITUD.
DR InterPro; IPR001060; -;
DR InterPro; IPR001452; -;
DR Pfam; PF00611; FCH; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00452; SH3DOMAIN.
DR PROSITE; PS00002; SH3; 1.

KW Mitosis; Cytoskeleton; SH3 domain; Phosphorylation; Coiled coil.
FT DOMAIN 24 110 FCH.
FT DOMAIN 108 207 COILED COIL (POTENTIAL).
FT DOMAIN 866 927 SH3.
SO SEQUENCE 927 AA; 102119 MW; FDCE7E0AAA3D247D CRC64;

Query Match 67.3%; Score 33; DB 1; Length 927;
Best Local Similarity 85.7%; Pred. No. 47;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KMDAEFR 10
Db 195 KMDAEFR 201

RESULT 14
BCPA_CHLLI STANDARD; PRT; 354 AA.
ID BCPA_CHLLI STANDARD; PRT; 354 AA.
AC Q46135;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE BACTERIOCHLOROPHYLL A PROTEIN (BCHL A PROTEIN)
DE OLSON PROTEIN (FMO-PROTEIN) (FRAGMENT).
GN FMOA.
OS Chlorobium limicola.
OC Bacteria; Green sulfur bacteria; Chlorobium.
OX NCBI_TaxID=1092;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95352646; PubMed=7626630;
RA Hager-Braun C., Xie D.L., Jarosch U., Herold E., Buttner M.,
RA Zimmermann R., Deutzmann R., Hauska G., Nelson N.;
RT "Stable photobleaching of P840 in Chlorobium reaction center
RT preparations: presence of the 42-kDa bacteriochlorophyll a protein
RT and a 17-kDa polypeptide.";
RL Biochemistry 34:9617-9624(1995).
CC -1- FUNCTION: INTERMEDIARY IN THE TRANSFER OF EXCITATION ENERGY FROM
CC THE CHLOROPHYLL TO THE REACTION CENTERS.
CC -1- SUBUNIT: HOMOTRIMER. EACH SUBUNIT CONTAINS 7 MOLECULES OF
CC BACTERIOCHLOROPHYLL A.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

EMBL; X83529; CAA58510.1; -;
DR HSSP; Q46393; 1KSA.
KW Electron transport; Photosynthesis; Reaction center; Magnesium;
KW Bacteriochlorophyll.
FT NON_TER 1
FT BINDING 99 99 BACTERIOCHLOROPHYLL A, 1 (BY SIMILARITY).
FT BINDING 134 134 BACTERIOCHLOROPHYLL A, 6 (BY SIMILARITY).
FT BINDING 278 278 BACTERIOCHLOROPHYLL A, 4 (BY SIMILARITY).
FT BINDING 285 285 BACTERIOCHLOROPHYLL A, 7 (BY SIMILARITY).
FT BINDING 286 286 BACTERIOCHLOROPHYLL A, 3 (BY SIMILARITY).
SO SEQUENCE 354 AA; 39243 MW; F4D4D565BDDCDB1B CRC64;

Query Match 65.3%; Score 32; DB 1; Length 354;
Best Local Similarity 75.0%; Pred. No. 27;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 EVKMDAEFR 9
Db 254 EVKVDGEF 261

```
RESULT 15
BCPA_CHLTE
ID BCPA_CHLTE STANDARD; PRT; 365 AA.
AC Q46393;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE BACTERIOCHLOROPHYLL A PROTEIN (BCHL A PROTEIN) (BCP) (FENNA-MATTHEWS-
DE OLSON PROTEIN) (FMO-PROTEIN).
DE FMOA.
GN Chlorobium tepidum.
OS Chlorobium tepidum.
OC Bacteria; Green sulfur bacteria; Chlorobium.
OX NCBI_TaxID=1097;
RN [1]
RP SEQUENCE FROM N.A.
RA Dracheva S., Williams J.A.C., Blankenship R.E.;
RT "Cloning and sequencing of the FMO-protein gene from Chlorobium
RT tepidum.";
RL (in) Murata N. (eds.);
RL Research in photosynthesis, pp.2:53-56, Kluwer Academic Publishers,
RL Dordrecht (1992).
RN [2]
RX X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE=97415773; PubMed=9268671;
RA Li Y.F., Zhou W., Blankenship R.E., Allen J.P.;
RT "Crystal structure of the bacteriochlorophyll a protein from
RT Chlorobium tepidum.";
RL J. Mol. Biol. 271:456-471(1997).
CC -!- FUNCTION: INTERMEDIARY IN THE TRANSFER OF EXCITATION ENERGY FROM
CC THE CHLOROPHYLL TO THE REACTION CENTERS.
CC -!- SUBUNIT: HOMOTRIMER. EACH SUBUNIT CONTAINS 7 MOLECULES OF
CC BACTERIOCHLOROPHYLL A.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L13700; AAA23111.1; -.
DR PDB; 1KSA; 25-FEB-98.
KW Electron transport; Photosynthesis; Reaction center; Magnesium;
KW 3D-structure; Bacteriochlorophyll.
FT INIT MET 0
FT BY SIMILARITY.
FT BINDING 110 110 BACTERIOCHLOROPHYLL A, 1.
FT BINDING 145 145 BACTERIOCHLOROPHYLL A, 6.
FT BINDING 289 289 BACTERIOCHLOROPHYLL A, 4.
FT BINDING 296 296 BACTERIOCHLOROPHYLL A, 7.
FT BINDING 297 297 BACTERIOCHLOROPHYLL A, 3.
SQ SEQUENCE 365 AA; 40163 MW; EB48DFE24DF6A780 CRC64;
Query Match 65.3%; Score 32; DB 1; Length 365;
Best Local Similarity 75.0%; Pred. No. 28;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 2 EVKMDAEF 9
DB 265 EVKVDGEF 272
|||:|
Search completed: September 6, 2001, 16:51:05
Job time: 809 sec
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:49:45 ; Search time 231.42 Seconds
(without alignments)
5.717 Million cell updates/sec

Title: US-09-603-713-4
Perfect score: 49
Sequence: 1 SEVKMDAEFR 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_16:*

- 1: sp-archaea:*
- 2: sp-bacteria:*
- 3: sp-fungi:*
- 4: sp-human:*
- 5: sp-invertebrate:*
- 6: sp-mammal:*
- 7: sp-mhc:*
- 8: sp-organelle:*
- 9: sp-phage:*
- 10: sp-plant:*
- 11: sp-rodent:*
- 12: sp-unclassified:*
- 13: sp-vertebrate:*
- 14: sp-virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	49	100.0	82	P78438	P78438 homo sapien
2	49	100.0	82	Q16014	Q16014 homo sapien
3	49	100.0	82	Q16019	Q16019 homo sapien
4	49	100.0	82	Q16020	Q16020 homo sapien
5	49	100.0	534	Q93296	Q93296 gallus gall
6	49	100.0	569	Q9PVL1	Q9PVL1 gallus gall
7	49	100.0	695	Q60496	Q60496 cavia porce
8	49	100.0	695	Q9DGJ8	Q9DGJ8 gallus gall
9	49	100.0	751	Q9DGJ7	Q9DGJ7 gallus gall
10	49	100.0	770	Q9TUI0	Q9TUI0 sus scrofa
11	44	89.8	79	Q35463	Q35463 cricetus
12	44	89.8	695	Q197487	Q197487 mus musculus
13	43	87.8	747	Q91963	Q91963 xenopus lae
14	42	85.7	423	Q45693	Q45693 burkholderi
15	42	85.7	423	Q52379	Q52379 pseudomonas
16	39	79.6	142	Q16896	Q16896 caenorhabd1
17	37	75.5	269	Q52512	Q52512 streptomyce
18	34	69.4	239	Q9FNC2	Q9FNC2 arabidopsis
19	34	69.4	605	Q9L1F6	Q9L1F6 streptomyce

20	34	69.4	1261	10	Q9LU30	Q9LU30 arabidopsis
21	34	69.4	3262	4	Q13788	Q13788 homo sapien
22	33	67.3	302	9	Q37840	Q37840 bacterioph
23	33	67.3	621	4	Q9H9V1	Q9H9V1 homo sapien
24	32	65.3	143	4	Q9H935	Q9H935 homo sapien
25	32	65.3	263	1	Q9HR12	Q9HR12 halobacteri
26	32	65.3	340	5	Q9U0X8	Q9U0X8 leishmania
27	32	65.3	376	14	Q9DVZ3	Q9DVZ3 pluteella xy
28	32	65.3	392	3	Q9P5G5	Q9P5G5 neurospora
29	32	65.3	426	1	Q9V2P8	Q9V2P8 pyrococcus
30	32	65.3	539	4	Q9NMA5	Q9NMA5 homo sapien
31	32	65.3	539	4	Q9NWD6	Q9NWD6 homo sapien
32	32	65.3	625	10	Q04086	Q04086 arabidopsis
33	32	65.3	700	2	Q9K679	Q9K679 bacillus ha
34	32	65.3	929	3	Q9HFI9	Q9HFI9 neurospora
35	32	65.3	1265	10	Q9LFO0	Q9LFO0 arabidopsis
36	32	65.3	2514	5	Q9Y061	Q9Y061 caenorhabd1
37	32	65.3	2531	5	Q22258	Q22258 caenorhabd1
38	31	63.3	155	1	Q9UZ40	Q9UZ40 pyrococcus
39	31	63.3	198	2	Q9K3T9	Q9K3T9 streptomyce
40	31	63.3	199	2	Q9Z7W4	Q9Z7W4 chlamydia p
41	31	63.3	226	1	Q26341	Q26341 methanobact
42	31	63.3	282	5	O02335	O02335 caenorhabd1
43	31	63.3	286	10	Q9SUP3	Q9SUP3 arabidopsis
44	31	63.3	345	4	Q9NPJ7	Q9NPJ7 homo sapien
45	31	63.3	346	4	Q9NPB6	Q9NPB6 homo sapien

ALIGNMENTS

RESULT 1

ID P78438 PRELIMINARY; PRT; 82 AA.
AC P78438;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE AMYLOID PROTEIN (BETA-AMYLOID PROTEIN) (FRAGMENT).
GN APP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89392030; PubMed=2675837;
RA Johnstone E.M., Chaney M.O., Moore R.E., Ward K.E., Norris F.H.,
RA Little S.P.;
RT "Alzheimer's disease amyloid peptide is encoded by two exons and shows
RT similarity to soybean trypsin inhibitor.";
RL Biochem. Biophys. Res. Commun. 163:1248-1255(1989).
RN [2]
RP SEQUENCE OF 19-48 FROM N.A.
RX MEDLINE=87120329; PubMed=2949367;
RA Tanzi R.E., Gusella J.F., Watkins P.C., Bruns G.A., George-Hyslop P.,
RA Van Keuren M.L., Patterson D., Pagan S., Kurnit D.M., Neve R.L.;
RT "Amyloid beta protein gene: cDNA, mRNA distribution, and genetic
RT linkage near the Alzheimer locus.";
RL Science 235:880-884(1987).
RN [3]
RP SEQUENCE OF 32-63 FROM N.A.
RX MEDLINE=93035397; PubMed=1415269;
RA Kamino K., Orr H.T., Payami H., Wijsman E.M., Alonso M.E., Pulst S.M.,
RA Anderson L., O'dahl S., Nemens E., White J.A.;
RT "Linkage and mutational analysis of familial Alzheimer disease
RT kindreds for the APP gene region.";
RL Am. J. Hum. Genet. 51:998-1014(1992).
DR EMBL; M29270; AAA51768.1; -;
DR EMBL; M29269; AAA51768.1; JOINED.
DR EMBL; M15532; AAA51564.1; -;
DR EMBL; S45136; AAB23646.1; -;
DR HSP; P05067; 1BA4.

```
FT NON_TER 1 1
SQ SEQUENCE 82 AA; 8994 MW; 8DA9E42B813A070E CRC64;

Query Match 100.0%; Score 49; DB 4; Length 82;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVKMDAEFR 10
Db 12 SEVKMDAEFR 21
|||||

RESULT 2
Q16014 PRELIMINARY; PRT; 82 AA.
AC Q16014;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE BETA-AMYL0ID PEPTIDE PRECURSOR (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93236601; PubMed=8476439;
RA Denman R.B.; Rosenzwaig R.; Miller D.L.;
RT "A system for studying the effect(s) of familial Alzheimer disease
RT mutations on the processing of the beta-amyloid peptide precursor.";
RL Biochem. Biophys. Res. Commun. 192:96-103(1993).
DR EMBL; S60721; AAB26263.2; -.
DR HSSP; P05067; 1BA4.
FT NON_TER 1
FT NON_TER 82
SQ SEQUENCE 82 AA; 8972 MW; F534AA5B3EA9230A CRC64;

Query Match 100.0%; Score 49; DB 4; Length 82;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVKMDAEFR 10
Db 13 SEVKMDAEFR 22
|||||

RESULT 3
Q16019 PRELIMINARY; PRT; 82 AA.
AC Q16019;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE BETA-AMYL0ID PEPTIDE PRECURSOR (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93236601; PubMed=8476439;
RA Denman R.B.; Rosenzwaig R.; Miller D.L.;
RT "A system for studying the effect(s) of familial Alzheimer disease
RT mutations on the processing of the beta-amyloid peptide precursor.";
RL Biochem. Biophys. Res. Commun. 192:96-103(1993).
DR EMBL; S61380; AAB26264.2; -.
DR HSSP; P05067; 1BA4.
FT NON_TER 1
FT NON_TER 82
SQ SEQUENCE 82 AA; 8938 MW; F534AA50E579230A CRC64;

Query Match 100.0%; Score 49; DB 4; Length 82;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVKMDAEFR 10
Db 13 SEVKMDAEFR 22
|||||

RESULT 4
Q16020 PRELIMINARY; PRT; 82 AA.
AC Q16020;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE BETA-AMYL0ID PEPTIDE PRECURSOR (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93236601; PubMed=8476439;
RA Denman R.B.; Rosenzwaig R.; Miller D.L.;
RT "A system for studying the effect(s) of familial Alzheimer disease
RT mutations on the processing of the beta-amyloid peptide precursor.";
RL Biochem. Biophys. Res. Commun. 192:96-103(1993).
DR EMBL; S61383; AAB26265.2; -.
DR HSSP; P05067; 1BA4.
FT NON_TER 1
FT NON_TER 82
SQ SEQUENCE 82 AA; 8882 MW; F534AA5AE5D9230A CRC64;

Query Match 100.0%; Score 49; DB 4; Length 82;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVKMDAEFR 10
Db 13 SEVKMDAEFR 22
|||||

RESULT 5
Q93296 PRELIMINARY; PRT; 534 AA.
AC Q93296;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE AMYL0ID PRECURSOR PROTEIN (FRAGMENT).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Barnes N.Y.; Ling L.; Yoshikawa K.; Schwartz L.M.; Oppenheim R.W.;
RA Milligan C.E.;
RT "Increased production of amyloid precursor protein provides a
RT substrate for Caspase 3 in dying motoneurons.";
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF042098; AAC25052.1; -.
DR HSSP; P05067; 1BA4.
DR InterPro; IPR001868; -.
DR PRINTS; PR00203; AMYL0IDA4.
DR PROSITE; PS00319; A4_EXTRA; 1.
DR PROSITE; PS00320; A4_INTRA; 1.
```



```
FT NON_TER 1 1
SQ SEQUENCE 534 AA; 60597 MW; FB53ECC2E66D4C92 CRC64;

Query Match 100.0%; Score 49; DB 13; Length 534;
Best Local Similarity 100.0%; Pred. No. 0.085;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SEVKMDAEFR 10
Db 431 SEVKMDAEFR 440

RESULT 6
Q9PVL1
ID Q9PVL1 PRELIMINARY; PRT; 569 AA.
AC Q9PVL1;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DE AMYLOID PROTEIN PRECURSOR (FRAGMENT).
GN APP.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRIN;
RA Coulson E.J., Paliga K., Beyreuther K., Masters C.L.;
RT "What the evolution of the amyloid protein precursor supergene family
RT tells us about its function.";
RL Neurochem. Int. 0:0-0(2000).
DR EMBL; AF030341; AAF12698.1; -.
DR HSSP; P05067; 1BA4.
DR InterPro; IPR001868; -.
DR Pfam; PF02177; A4_EXTRA; 1.
DR PRINTS; PR00203; AMYLOIDA4.
DR PROSITE; PS00319; A4_EXTRA; 1.
DR PROSITE; PS00320; A4_INTRA; 1.
DR SMART; SM00006; A4_EXTRA; 1.
FT NON_TER 1 1
SQ SEQUENCE 569 AA; 64753 MW; 0AB8BB851863A19D CRC64;

Query Match 100.0%; Score 49; DB 13; Length 569;
Best Local Similarity 100.0%; Pred. No. 0.091;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SEVKMDAEFR 10
Db 467 SEVKMDAEFR 476

RESULT 7
Q60496
ID Q60496 PRELIMINARY; PRT; 695 AA.
AC Q60496;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
DE ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystriognathii; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRIN;
RA Beck M., Mueller D., Bigl V.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: FUNCTIONAL RECEPTOR WHICH COUPLES TO
```

```
CC INTRACELLULAR SIGNALING PATHWAY THROUGH THE GTP-BINDING PROTEIN
CC G(O).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- SIMILARITY: TO OTHER SPECIES APP ANALOGUES.
DR EMBL; X97631; CAA66230.1; -.
DR HSSP; P05067; 1BA4.
DR InterPro; IPR001868; -.
DR Pfam; PF02177; A4_EXTRA; 1.
DR PRINTS; PR00203; AMYLOIDA4.
DR PROSITE; PS00319; A4_EXTRA; 1.
DR PROSITE; PS00320; A4_INTRA; 1.
DR SMART; SM00006; A4_EXTRA; 1.
SQ SEQUENCE 695 AA; 78701 MW; 5196A0C4017F16AB CRC64;

Query Match 100.0%; Score 49; DB 11; Length 695;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SEVKMDAEFR 10
Db 592 SEVKMDAEFR 601

RESULT 8
Q9DGJ8
ID Q9DGJ8 PRELIMINARY; PRT; 695 AA.
AC Q9DGJ8;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
DE BETA-AMYLOID PRECURSOR PROTEIN 695 ISOFORM.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Sarasa M., Rodolosse A., Sorribas V.;
RT "Cloning of full-length chicken beta-amyloid precursor protein
RT isoforms.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF285218; AAG00593.1; -.
SQ SEQUENCE 695 AA; 78565 MW; F201ED02AEC86D95 CRC64;

Query Match 100.0%; Score 49; DB 13; Length 695;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SEVKMDAEFR 10
Db 592 SEVKMDAEFR 601

RESULT 9
Q9DGJ7
ID Q9DGJ7 PRELIMINARY; PRT; 751 AA.
AC Q9DGJ7;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
DE BETA-AMYLOID PRECURSOR PROTEIN 751 ISOFORM.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Sarasa M., Rodolosse A., Sorribas V.;
RT "Cloning of full-length chicken beta-amyloid precursor protein
```

```

RT isoforms."
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RR EMBL; AF289219; AAG00594.1; -.
SQ SEQUENCE 751 AA; 84705 MW; E78E9413A8033D84 CRC64;

Query Match 100.0%; Score 49; DB 13; Length 751;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 10; Conservative 0; Mismatches 0; Indels 0;

Qy 1 SEVKMDAEFR 10
   |||||
Db 648 SEVKMDAEFR 657

RESULT 10
Q9TU10 PRELIMINARY; PRT; 770 AA.
ID ID Q9TU10
AC AC Q9TU10
DT DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DT DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE DE AMYLOID PRECURSOR PROTEIN.
OS OS Sus scrofa (Pig).
OC OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX OX NCBI_TaxID=9823;
RN RN [1]
RP RP SEQUENCE FROM N.A.
RA RA Kimura A., Takahashi T.;
RT RT "Amyloid Precursor Protein 770.";
RL RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC CC -!- SIMILARITY: TO PANCREATIC TRYPsin INHIBITOR (KUNITZ) DOMAIN.
DR DR EMBL; AB032550; BAA84580.1; -.
DR DR HSSP; P05067; 1AAP.
DR DR InterPro; IPR001868; -.
DR DR InterPro; IPR002223; -.
DR DR Pfam; PF00014; Kunitz_BPTI; 1.
DR DR Pfam; PF02177; A4_EXTRA; 1.
DR DR PRINTS; PR00203; AMYLOIDA4.
DR DR PRINTS; PR00759; BASICPTASE.
DR DR PROSITE; PS00319; A4_EXTRA; 1.
DR DR PROSITE; PS00320; A4_INTRA; 1.
DR DR PROSITE; PS00280; BPTI_KUNITZ; 1.
DR DR SMART; SM00006; A4_EXTRA; 1.
DR DR Serine protease inhibitor.
SQ SEQUENCE 770 AA; 86961 MW; 5F7ALDCB2BC583E CRC64;

```

Query Match 100.0%; Score 49; DB 6; Length 770;
Best Local Similarity 100.0%; Pred. NO. 0.12;
10. Conserved 0: Mismatches 0: Indels 0: Gaps 0:

QY	1	SEVKMDAEFR	10
Db	667	SEVKMDAEFR	676

RESULT	11.
ID	O35463
AC	PRELIMINARY;
DC	PRT; 79 AA.
DT	01-JAN-1998 (TReMBLrel. 05, Created)
DT	01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT	01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE	DE ALZHEIMER'S AMYLOID BETA PROTEIN PRECURSOR (FRAGMENT).
GN	BETA APP.
OS	Cricetulus griseus (Chinese hamster).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC	Cricetulus.
RN	[1]

RP	SEQUENCE FROM N.A.	
RA	Sambamurti K., Plinnix I., Gandhi S.;	
FL	Submitted (Oct-1997) to the EMBL/GenBank/DBJ databases.	
FL	EMBL: AF030413; AAB86608.1; -.	
DR	HSP: P05067; IQCM.	
FT	NON_TER	1
FT	NON_TER	79
FT	NON_TER	79
SO	SEQUENCE	79 AA; 8538 MW; 37FC26C3BFF3F597 CRC64;

Query Match	89.8%;	Score 44;	DB 11;	Length 79;
Best Local Similarity	100.0%;	Pred. No. 0.12;		
Matches	9;	Conservative 0;	Mismatches 0;	Indels 0;

QY 1 SEVKMDAEF 9
Db 16 SEVKMDAEF 24

RESULT 12		
P97487	PRELIMINARY;	PRT; 695 AA.
ID	P97487	
AC	P97487; P97942;	
DT	01-MAY-1997 (TReMBLrel. 03, Created)	
DC	01-MAY-1997 (TReMBLrel. 03, Last sequence update)	
DT	01-MAR-2001 (TReMBLrel. 16, Last annotation update)	
DE	HIPOCAMPAL AMYLOID PROTEIN.	
DE	APP.	
GN	Mus musculus (Mouse).	
OS	Amphibia; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
OX	NCBI_Taxid=10090;	
OX	(1)	
RN	SEQUENCE FROM N.A.	
RP	STRAIN=SAMP8; TISSUE=HIPPOCAMPUS;	
RC	Flood J.F., Kumar V.B., Sasser T., Word I., Morley J.E.;	
EA	Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.	
DT		

1	SEQUENCE OF 581-662 FROM N.A.	
2	STRAIN=129SV;	
3	WRAG M.A., Busfield F., Duff K., Korenblat K., Capechi M.,	
4	Loring J.F., Goate A.M.;	
5	Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.	
6	EMBL: U84012; AAB41502.1; -;	
7	EMBL: U82624; AAB40919.1; -;	
8	HSSP: P05067; 1QCM.	
9	InterPro: IPR001868; -;	
10	Pfam: PF02177; A4_EXTRA; 1.	
11	PRINTS: PR00203; AMYLOIDA4.	
12	PROSITE: PS00319; A4_EXTRA; 1.	
13	PROSITE: PS00320; A4_INTRA; 1.	
14	SMART: SM00006; A4_EXTRA; 1.	
15	SEQUENCE 695 AA; 78414 MW; 9A5FB2ED261236E CRC64;	

Query Match	89.8%;	Score 44;	DB 11;	Length 695;
Best Local Similarity	100.0%;	Pred. No. 1.2;		
Matches	9: Conservative	0: Mismatches	0: Indels	0: Caps

QY	1	SEVKMDAEF	9
Dh	592	SEVKMDAEF	600

RESULT	13
Q91963	
ID	PRELIMINARY; PRT; 747 AA.
AC	Q91963;
DT	01-NOV-1996 (TRENBLrel. 01, Created)
DT	01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT	01-MAR-2001 (TRENBLrel. 16, Last annotation update)
DE	APP747.
OS	Xenopus laevis (African clawed frog).

DE APP/4/.
OS *Xenopus laevis* (African clawed frog).

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
RN NCBI_TaxID=8355;
RX [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-93129227; PubMed=1282805;
RA Okada H., Okamoto H.;
RT "A Xenopus homologue of the human beta-amyloid precursor protein:
RT developmental regulation of its gene expression.";
RL Biochem. Biophys. Res. Commun. 189:1561-1568(1992).
CC -1- SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
DR EMBL: S52417; AAB24853.1; -.
DR HSP; P05067; IQCM.
DR InterPro; IPR001868; -.
DR InterPro; IPR002223; -.
DR Pfam; PF00014; Kunitz_BPTI; 1.
DR Pfam; PF02177; A4_EXTRA; 1.
DR PRINTS; PR00203; AMYLOIDAM.
DR PRINTS; PR00759; BASICPTASE.
DR PROSITE; PS00319; A4_EXTRA; 1.
DR PROSITE; PS00320; A4_INTRA; 1.
DR PROSITE; PS00280; BPTI_KUNITZ; 1.
DR SMART; SM00131; KU; 1.
KW Serine protease inhibitor.
SQ SEQUENCE 747 AA; 84892 MW; A75E81885681D948 CRC64;

Query Match 87.8%; Score 43; DB 13; Length 747;
Best Local Similarity 80.0%; Pred. No. 2;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVKMDAEFR 10
DB 644 SEVKMDSEYR 653
|||||:|:|

RESULT 14
Q45693 PRELIMINARY; PRT; 423 AA.
AC Q45693;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TRENBLrel. 13, Last annotation update)
DE FERREDOXIN OXIDOREDUCTASE (DNTAA), ORF2 PROTEIN, FERREDOXIN (DNTAB),
DE ISP-ALPHA (DNTAC), ISP-BETA (DNTAD) GENES, COMPLETE CDS (DNTAB)
DE (DNTAD).
OS Burkholderia sp.
OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
OC Burkholderia.
OX NCBI_TaxID=36773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DNT;
RA Suen W.C., Haigler B.E., Spain J.C.;
RL J. Bacteriol. 178:0-0(0).
DR EMBL; U62430; AAB09764.1; -.
DR InterPro; IPR001281; -.
DR InterPro; IPR001663; -.
DR Pfam; PF00355; Rieske; 1.
DR Pfam; PF00848; Ring_hydroxyl_A; 1.
DR PROSITE; PS00570; RING_HYDROXYL_ALPHA; UNKNOWN_1.
SQ SEQUENCE 423 AA; 48916 MW; 81B346D7A2B4490E CRC64;

Query Match 85.7%; Score 42; DB 2; Length 423;
Best Local Similarity 80.0%; Pred. No. 1.8;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVKMDAEFR 10
DB 244 SELKMDAEFR 253
|||||:|:|

us-09-603-713-4.rspt

RESULT 15
Q52379 PRELIMINARY; PRT; 423 AA.
AC Q52379;
DT 01-JUN-1998 (TRENBLrel. 06, Created)
DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
DT 01-MAY-2000 (TRENBLrel. 13, Last annotation update)
DE SALICYLATE-5-HYDROXYLASE LARGE OXYGENASE COMPONENT.
GN NAGG.
OS Pseudomonas sp.
OG Plasmid pMWU2.
OC Bacteria; Proteobacteria.
OX NCBI_TaxID=306;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-U2;
RX MEDLINE-9833751; PubMed=9573207;
RA Fuenmayor S., Wild M., Boyes A.L., Williams P.A.;
RT "A gene cluster encoding steps in conversion of naphthalene to
RT gentisate in Pseudomonas sp. strain U2.";
RL J. Bacteriol. 180:2522-2530(1998).
DR EMBL; AF036940; AAD12607.1; -.
DR InterPro; IPR001281; -.
DR InterPro; IPR001663; -.
DR Pfam; PF00355; Rieske; 1.
DR Pfam; PF00848; Ring_hydroxyl_A; 1.
DR PRINTS; PR00090; RING_HYDROXYL_ALPHA; UNKNOWN_1.
DR PROSITE; PS00570; RING_HYDROXYL_ALPHA; UNKNOWN_1.
KW Plasmid.
SQ SEQUENCE 423 AA; 48795 MW; 2EDC17B117974DC2 CRC64;

Query Match 85.7%; Score 42; DB 2; Length 423;
Best Local Similarity 80.0%; Pred. No. 1.8;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVKMDAEFR 10
DB 244 SELKMDAEFR 253
|||||:|:|

Search completed: September 6, 2001, 16:49:46
Job time: 730 sec
```


GenCore version 4.5
 Copyright (c) 1993 - 2000 CompuGen Ltd.
 OM protein - protein search, using sw model
 Run on: September 6, 2001, 16:43:27 ; Search time 225.25 Seconds
 (without alignments)
 2.691 Million cell updates/sec

Title: US-09-603-713-5
 Perfect score: 49
 Sequence: 1 SEVNLDAEFR 10
 Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5
 Searched: 412676 seqs, 60623988 residues
 Total number of hits satisfying chosen parameters: 412676
 Minimum DB seq length: 0
 Maximum DB seq length: 2000000000
 Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : A_Geneseq_0601.*
 1: /SID8/gcgdata/geneseq/geneseq/AA1980.DAT.*
 2: /SID8/gcgdata/geneseq/geneseq/AA1981.DAT.*
 3: /SID8/gcgdata/geneseq/geneseq/AA1982.DAT.*
 4: /SID8/gcgdata/geneseq/geneseq/AA1983.DAT.*
 5: /SID8/gcgdata/geneseq/geneseq/AA1984.DAT.*
 6: /SID8/gcgdata/geneseq/geneseq/AA1985.DAT.*
 7: /SID8/gcgdata/geneseq/geneseq/AA1986.DAT.*
 8: /SID8/gcgdata/geneseq/geneseq/AA1987.DAT.*
 9: /SID8/gcgdata/geneseq/geneseq/AA1988.DAT.*
 10: /SID8/gcgdata/geneseq/geneseq/AA1989.DAT.*
 11: /SID8/gcgdata/geneseq/geneseq/AA1990.DAT.*
 12: /SID8/gcgdata/geneseq/geneseq/AA1991.DAT.*
 13: /SID8/gcgdata/geneseq/geneseq/AA1992.DAT.*
 14: /SID8/gcgdata/geneseq/geneseq/AA1993.DAT.*
 15: /SID8/gcgdata/geneseq/geneseq/AA1994.DAT.*
 16: /SID8/gcgdata/geneseq/geneseq/AA1995.DAT.*
 17: /SID8/gcgdata/geneseq/geneseq/AA1996.DAT.*
 18: /SID8/gcgdata/geneseq/geneseq/AA1997.DAT.*
 19: /SID8/gcgdata/geneseq/geneseq/AA1998.DAT.*
 20: /SID8/gcgdata/geneseq/geneseq/AA1999.DAT.*
 21: /SID8/gcgdata/geneseq/geneseq/AA2000.DAT.*
 22: /SID8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49	100.0	10	AAW08362	Beta-secretase sub
2	49	100.0	10	AAW08362	Synthetic oligopep
3	49	100.0	10	AAW08362	Beta-APP alpha-sec
4	49	100.0	10	AAW08362	Synthetic peptide
5	49	100.0	10	AAW08362	Swedish mutation p
6	49	100.0	16	AAW08362	Human beta-amyloid
7	49	100.0	20	AAW08362	Beta-APP alpha-sec
8	49	100.0	21	AAW08362	Beta-secretase sub
9	49	100.0	21	AAW08362	Synthetic oligopep
10	49	100.0	32	AAW08362	Mouse amyloid prec
11	49	100.0	33	AAW08362	Beta-secretase sub

12	49	100.0	33	20	AAW08362	Synthetic oligopep
13	49	100.0	39	21	AAW08362	Beta-APP alpha-sec
14	49	100.0	42	18	AAW08362	Wild type APP beta
15	49	100.0	42	20	AAW08362	Synthetic oligopep
16	49	100.0	58	20	AAW08362	Swedish-FAD APP714
17	49	100.0	115	20	AAW08362	Swedish-FAD APP po
18	49	100.0	115	20	AAW08362	Swedish-FAD APP po
19	49	100.0	506	19	AAW08362	Maltose binding pr
20	49	100.0	506	20	AAW08362	MBP-APP (SW192) fu
21	49	100.0	695	21	AAW08362	Human APP695-sw va
22	49	100.0	697	21	AAW08362	Human APPSW-KK anl
23	47	95.9	10	21	AAW08362	Beta-APP alpha-sec
24	47	95.9	20	21	AAW08362	Beta-APP alpha-sec
25	45	91.8	103	16	AAW08362	Beta-amyloid precu
26	44	89.8	9	19	AAW08362	Fluorogenic protea
27	44	89.8	9	21	AAW08362	A peptide fragment
28	44	89.8	9	21	AAW08362	Substrate for beta
29	44	89.8	20	19	AAW08362	Fluorogenic protea
30	44	89.8	21	19	AAW08362	Fluorogenic protea
31	44	89.8	30	18	AAW08362	Beta-secretase sub
32	44	89.8	30	20	AAW08362	Synthetic oligopep
33	44	89.8	30	21	AAW08362	Substrate for beta
34	44	89.8	32	17	AAW08362	Mouse amyloid prec
35	44	89.8	32	17	AAW08362	Mouse amyloid prec
36	44	89.8	33	21	AAW08362	Substrate for beta
37	44	89.8	695	18	AAW08362	APP695 mutant A-be
38	44	89.8	695	18	AAW08362	APP695 mutant A-be
39	44	89.8	751	18	AAW08362	APP751 mutant A-be
40	44	89.8	751	18	AAW08362	APP751 mutant A-be
41	44	89.8	770	18	AAW08362	APP770 mutant A-be
42	44	89.8	770	18	AAW08362	APP770 mutant A-be
43	43	87.8	10	21	AAW08362	Beta-APP alpha-sec
44	43	87.8	10	21	AAW08362	Beta-APP alpha-sec
45	43	87.8	20	21	AAW08362	Beta-APP alpha-sec

ALIGNMENTS

RESULT 1	
AAW08362	
ID AAW08362 standard; peptide; 10 AA.	
XX	
AC AAW08362;	
XX	
DT 05-SEP-1997 (first entry)	
XX	
DE Beta-secretase substrate #3.	
XX	
KW Beta-cleavage site; beta amyloid precursor protein; APP; beta-secretase;	
KW alpha-secretase; proteolytic cleavage; inhibitor; Alzheimer's disease.	
XX	
OS Synthetic.	
XX	
FH Key Location/Qualifiers	
FT Misc-difference 1	
FT	/note= "acetylated"
XX	
PN WO9640885-A2.	
XX	
PD 19-DEC-1996.	
XX	
PF 07-JUN-1996; 96WO-0509985.	
XX	
PR 07-JUN-1995; 95US-0485152.	
PR	07-JUN-1995; 95US-0480498.
XX	
PA (ATHE-) ATHENA NEUROSCIENCES INC.	
XX	
PI Anderson JP, Chrysler SMS, Jacobson-croak KL, Keim PS;	
PI Mcconlogue LC, Sinha S, Tan H;	
XX	
DR WPI; 1997-052304/05.	

XX Beta-secretase which specifically cleaves beta-amyloid precursor
 PT protein - useful to screen for inhibitors useful in treatment of
 PT Alzheimer's disease
 XX
 PS Disclosure; Page 45; 92pp; English.
 CC
 CC AAW08359-W08362 represent substrates for the enzyme of the invention.
 CC The enzyme of the invention is beta-secretase, and specifically cleaves
 CC beta-amyloid precursor protein (beta-APP). Normal processing of beta-APP
 CC is thought to occur via cleavage between residues 16 and 17 of the
 CC beta-amyloid peptide region by an alpha-secretase. Pathogenic processing
 CC is thought to occur by beta-secretase cleavage of beta-APP.
 CC Beta-secretase activity can be detected and measured using a method of
 CC the invention, which detects at least one of the beta-secretase cleavage
 CC products formed on cleavage. The method can be used to determine whether
 CC a test substance inhibits proteolytic cleavage, by beta-secretase, of
 CC beta-APP. Compounds effective to at least partially inhibit
 CC beta-secretase activity can be used to inhibit cleavage of beta-APP in
 CC cells or mammalian hosts. Isolation and purification of beta-secretase
 CC will permit chemical modelling of a critical event in the pathology of
 CC Alzheimer's disease.
 XX
 SQ Sequence 10 AA;

Query Match 100.0%; Score 49; DB 18; Length 10;
 Best Local Similarity 100.0%; Pred. No. 0.00056;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SEVNLDAEFR 10
 |||||
 DB 1 sevnldaefr 10

RESULT 2

AAV33756
 ID AAY33756 standard; Protein; 10 AA.

AC AAY33756;

XX 09-NOV-1999 (first entry)

XX Synthetic oligopeptide 5-5'SW.

XX Beta-secretase; beta-amyloid protein precursor; APP; Down's syndrome;
 KW Alzheimer's disease; measure activity; cleavage site.

XX Synthetic.

XX Key Location/Qualifiers
 FH Modified-site 1 /note= "N-terminal Ser is acetylated"
 FT

XX US5942400-A.

XX 24-AUG-1999.

XX 07-JUN-1996; 96US-0659984.

XX 07-JUN-1996; 96US-0659984.

XX 07-JUN-1995; 95US-0480498.

XX 07-JUN-1995; 95US-0485152.

XX (ELAN-) ELAN PHARM INC.

XX Anderson JP, Jacobson-Croak KL, Sinha S;

XX WPI; 1999-517417/43.

XX A method for detecting human beta-secretase cleavage of polypeptides

XX useful for identifying beta-secretase inhibitors

XX

PS Examples; Column 30; 43pp; English.

XX Sequences AAY33752-Y33756 are synthetic oligopeptides used for measuring
 CC the activity of beta-secretase (AAY33741). Beta-secretase is capable of
 CC cleaving beta-amyloid protein precursor (APP) (AAY33742). These
 CC synthetic peptides contain the cleavage site of APP. Beta-secretase and
 CC APP are used in a method for detecting human beta-secretase cleavage of
 CC polypeptides and for identifying beta-secretase inhibitors. Inhibition
 CC of beta-secretase activity would be useful for chemical modelling of a
 CC critical event in the pathology of Alzheimer's disease. Inhibitors of
 CC beta-secretase would be useful for the prevention and treatment of
 CC Alzheimer's disease and Down's Syndrome.

XX Sequence 10 AA;

Query Match 100.0%; Score 49; DB 20; Length 10;
 Best Local Similarity 100.0%; Pred. No. 0.00056;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SEVNLDAEFR 10
 |||||
 DB 1 sevnldaefr 10

RESULT 3

AAV69707

ID AAY69707 standard; peptide; 10 AA.

XX AAY69707;

XX 11-APR-2000 (first entry)

XX Beta-APP alpha-secretase substrate [NLD]-APP(-5,+5).

KW Nootropic; neuroprotective; beta-amyloid precursor protein; metabolism;
 KW cleavage site; beta-secretase; neurodegenerative disease;
 KW Alzheimer's disease.

XX Homo sapiens.

XX Synthetic.

XX WO9964587-Al.

XX 16-DEC-1999.

XX 04-JUN-1999; 99WO-FR01326.

XX 05-JUN-1998; 98FR-0007068.

XX 31-MAR-1999; 99US-0122599.

XX (RHON) RHONE-POULENC RORER SA.

XX (UYPA-) UNIV CURIE PARIS VI P & M.

XX Rholam M, Munoz-Gimenez N, Moutaouakil M, Cohen P, Bertrand P;

XX WPI; 2000-097537/08.

XX Polypeptide with beta-secretase activity, specific for wild-type

XX amyloid precursor protein, useful in treating Alzheimer's disease

XX Example 3; Page 24; 44pp; French.

XX Peptides AAY69702-Y69718 represent synthetic peptide substrates for a
 CC novel polypeptide with beta-secretase activity that can cleave
 CC specifically the natural beta-amyloid precursor protein (bAPP). Normal
 CC cleavage of the protein occurs between amino acids Met596-Asp597 and
 CC Val636-Ile637 (positions 4-5 and 44-45 of AAY69701). The novel
 CC polypeptide is used to identify agents that interact specifically with
 CC it. These agents regulate metabolism of APP, particularly they slow down
 CC or reduce production of beta-amyloid, so can be used to treat
 CC neurodegenerative diseases, particularly Alzheimer's disease.

XX

SQ Sequence 10 AA;

Query Match 100.0%; Score 49; DB 21; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.00056;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVNLDAEFR 10
| | | | | | | | | |
Db 1 sevnldaefr 10

RESULT 4

AAB66575
ID AAB66575 standard; Peptide; 10 AA.

XX AC AAB66575;

XX DT 12-APR-2001 (first entry)

XX DE Synthetic peptide derived from APP beta-secretase site.

XX KW Memapsin 2; neurotropic; neuroprotective; amyloid precursor protein;

XX KW APP; memapsin 2 inhibitor; Alzheimer's disease.

XX OS Synthetic.

XX PN WO200100665-A2.

XX PD 04-JAN-2001.

XX PF 27-JUN-2000; 2000WO-US17742.

XX PR 28-JUN-1999; 99US-0141363.

XX PR 30-NOV-1999; 99US-0168060.

XX PR 25-JAN-2000; 2000US-0177836.

XX PR 27-JAN-2000; 2000US-0178368.

XX PR 08-JUN-2000; 2000US-0210292.

XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.

XX PA (UNIV) UNIV ILLINOIS FOUND.

XX PI Tang JJN, Hong L, Ghosh AK;

XX WPI; 2001-137933/14.

XX Novel memapsin 2 inhibitors which bind to active site of memapsin 2
PT having 2 catalytic aspartic residues and substrate binding cleft, used
PT to treat Alzheimer's disease by blocking amyloid precursor protein
PT cleavage -

XX PS Disclosure; Page 11; 86pp; English.

XX CC The present sequence is given in a specification relating to an inhibitor
CC of catalytically active memapsin 2. The inhibitor binds to the memapsin 2
CC active site, which is defined by the presence of two catalytic aspartic
CC residues and a substrate binding cleft. The inhibitor is useful for
CC the treatment and diagnosis of Alzheimer's disease. It is useful in
CC screens for individuals with a genetic predisposition to Alzheimer's
CC disease. The inhibitor is useful as a reagent for specifically binding to
CC memapsin 2 or memapsin 2 analogues and for aiding in memapsin 2
CC isolation, purification and characterisation.

XX SQ Sequence 10 AA;

Query Match 100.0%; Score 49; DB 22; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.00056;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVNLDAEFR 10

| | | | | | | | | |
Db 1 sevnldaefr 10

RESULT 5

AAB61337

ID AAB61337 standard; peptide; 10 AA.

XX AC AAB61337;

XX DT 02-APR-2001 (first entry)

XX DE Swedish mutation peptide from beta amyloid precursor protein.

XX KW Memapsin 2; catalyst; Alzheimer's.

XX OS Unidentified.

XX PN WO200100663-A2.

XX PD 04-JAN-2001.

XX PF 27-JUN-2000; 2000WO-US17661.

XX PR 28-JUN-1999; 99US-0141363.

XX PR 30-NOV-1999; 99US-0168060.

XX PR 25-JAN-2000; 2000US-0177836.

XX PR 27-JAN-2000; 2000US-0178368.

XX PR 08-JUN-2000; 2000US-0210292.

XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.

XX PI Tang JJN, Lin X, Koelsch G;

XX WPI; 2001-102885/11.

XX PT Purified recombinant catalytically active memapsin 2, used to screen
PT inhibitors of it, which are used to treat and prevent Alzheimer's
PT disease -

XX PS Claim 6; Page 11; 86pp; English.

XX CC The present invention relates to a purified recombinant
CC catalytically active memapsin 2. The invention may be used for
CC isolating inhibitors which are used to treat or prevent
CC Alzheimer's disease. The invention may also be used to screen
CC for individuals more genetically prone to develop Alzheimer's
CC disease.

XX SQ Sequence 10 AA;

Query Match 100.0%; Score 49; DB 22; Length 10;

Best Local Similarity 100.0%; Pred. No. 0.00056;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVNLDAEFR 10

| | | | | | | | | |
Db 1 sevnldaefr 10

RESULT 6

AAB06316

ID AAB06316 standard; peptide; 16 AA.

XX AC AAB06316;

XX DT 03-OCT-2000 (first entry)

XX DE Human beta-amyloid precursor protein beta-secretase site.

XX KW Human; beta-amyloid precursor protein; beta-APP; beta-secretase;

XX KW subtilisin-kexin isoenzyme 1; SKI-1;

XX KW pro-brain-derived neurotrophic factor; proBDNF; antilipemia;

XX KW cytostatic; vasotropic; SKI-1 inhibitor; hypercholesterolaemia;

KW liver steatosis; Ras-dependent cancer; restenosis;
 KW amyloid protein formation.
 XX Mammalia.

XX Key Location/Qualifiers
 FT Cleavage-site 8..9
 XX WO200026348-A2.

XX PD 11-MAY-2000.
 XX 04-NOV-1999; 99WO-CA01058.
 XX 04-NOV-1998; 98CA-2249648.

XX (RECL-) INST RECH CLINIQUES MONTREAL.
 XX Seidah N, Chretien M, Marcinkiewicz M, Laaksonen R, Davignon J;
 XX WPI; 2000-365601/31.

XX Novel soluble proteic fragment of subtilisin-kexin isoenzyme for
 PT producing a polypeptide useful for treating hypercholesterolemia, liver
 PT steatosis and amyloidosis, comprises a specific amino acid sequence
 XX Example 4; Page 51; 119pp; English.

XX The present sequence is the beta-secretase site of human beta-amyloid
 CC precursor protein (beta-APP). The sequence may be cleaved by a mammalian
 CC subtilase, specifically subtilisin-kexin isoenzyme 1 (SKI-1), a type-1
 CC membrane-bound proteinase. Peptides which bind to and are cleaved by
 CC of SKI-1 activity, or for screening enhancers of SKI-1 activity. Proteic
 CC fragments of SKI-1 which bind to the SKI-1 catalytic site may be used as
 CC inhibitors of SKI-1 activity. They may be used to treat diseases
 CC involving overexpression of SKI-1 or SKI-1 substrate. Such diseases
 CC include hypercholesterolemia, high levels of fatty acids, lipids or
 CC farnesyl pyrophosphate, liver steatosis, Ras-dependent cancer, restenosis
 CC and amyloid protein formation.

XX Sequence 16 AA;

Query Match 100.0%; Score 49; DB 21; Length 16;
 Best Local Similarity 100.0%; Pred. No. 0.00097;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVNLDAEFR 10
 |||||
 Db 4 sevnldaefr 13

RESULT 7
 AAY69714
 ID AAY69714 standard; peptide; 20 AA.
 XX AC AAY69714;
 XX 11-APR-2000 (first entry)

XX Beta-APP alpha-secretase substrate [NL]-APP(-10,+10).
 XX Neurotrophic; neuroprotective; beta-amyloid precursor protein; metabolism;
 KW cleavage site; beta-secretase; neurodegenerative disease;
 KW Alzheimer's disease.
 XX Homo sapiens.
 OS Synthetic.
 XX WO9964587-A1.
 XX 16-DEC-1999.

XX 04-JUN-1999; 99WO-FR01326.
 XX 05-JUN-1998; 98FR-0007068.
 XX 31-MAR-1999; 99US-0122599.
 XX (RHON) RHONE-POULENC RORER SA.
 XX (UYPA-) UNIV CURIE PARIS VI P & M.
 XX Rholam M, Munoz-Gimenez N, Moutaouakil M, Cohen P, Bertrand P;
 XX WPI; 2000-097537/08.

XX Polypeptide with beta-secretase activity, specific for wild-type
 PT amyloid precursor protein, useful in treating Alzheimer's disease
 XX Example 3; Page 24; 44pp; French.

XX Peptides AAY69702-Y69718 represent synthetic peptide substrates for a
 CC novel polypeptide with beta-secretase activity that can cleave
 CC specifically the natural beta-amyloid precursor protein (BAPP). Normal
 CC cleavage of the protein occurs between amino acids Met596-Asp597 and
 CC Val636-Ile637 (positions 4-5 and 44-45 of AAY69701). The novel
 CC polypeptide is used to identify agents that interact specifically with
 CC it. These agents regulate metabolism of APP, particularly they slow down
 CC or reduce production of beta-amyloid, so can be used to treat
 CC neurodegenerative diseases, particularly Alzheimer's disease.

XX Sequence 20 AA;

Query Match 100.0%; Score 49; DB 21; Length 20;
 Best Local Similarity 100.0%; Pred. No. 0.0013;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVNLDAEFR 10
 |||||
 Db 6 sevnldaefr 15

RESULT 8
 AAW08361
 ID AAW08361 standard; peptide; 21 AA.
 XX AC AAW08361;
 XX 05-SEP-1997 (first entry)

XX Beta-secretase substrate #3.
 XX Beta-cleavage site; beta amyloid precursor protein; APP; beta-secretase;
 KW alpha-secretase; proteolytic cleavage; inhibitor; Alzheimer's disease.
 XX Synthetic.
 XX WO9640885-A2.
 XX 19-DEC-1996.
 XX 07-JUN-1996; 96WO-US09985.
 XX 07-JUN-1995; 95US-0485152.
 XX 07-JUN-1995; 95US-0480498.
 XX (ATHE-) ATHENA NEUROSCIENCES INC.
 XX Anderson JP, Chrysler SMS, Jacobson-croak KL, Keim PS;
 XX Mcconlogue LC, Sinha S, Tan H;
 XX WPI; 1997-052304/05.

XX Beta-secretase which specifically cleaves beta-amyloid precursor
 PT protein - useful to screen for inhibitors useful in treatment of

PT Alzheimer's disease
 XX Disclosure; Page 45; 92pp; English.
 CC AAW08359-W08362 represent substrates for the enzyme of the invention.
 CC The enzyme of the invention is beta-secretase, and specifically cleaves
 CC beta-amyloid precursor protein (beta-APP). Normal processing of beta-APP
 CC is thought to occur via cleavage between residues 16 and 17 of the
 CC beta-amyloid peptide region by an alpha-secretase. Pathogenic processing
 CC is thought to occur by beta-secretase cleavage of beta-APP.
 CC Beta-secretase activity can be detected and measured using a method of
 CC the invention, which detects at least one of the beta-secretase cleavage
 CC products formed on cleavage. The method can be used to determine whether
 CC a test substance inhibits proteolytic cleavage, by beta-secretase, of
 CC beta-APP. Compounds effective to at least partially inhibit
 CC beta-secretase activity can be used to inhibit cleavage of beta-APP in
 CC cells or mammalian hosts. Isolation and purification of beta-secretase
 CC will permit chemical modelling of a critical event in the pathology of
 CC Alzheimer's disease.
 XX
 SQ Sequence 21 AA;
 Query Match 100.0%; Score 49; DB 18; Length 21;
 Best Local Similarity 100.0%; Pred. No. 0.0013;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SEVNLDAEFR 10
 DB 1 sevnlदैfr 10
 |||||
 RESULT 9
 AAY33755
 ID AAY33755 standard; Protein; 21 AA.
 XX
 AC AAY33755;
 XX
 DT 09-NOV-1999 (first entry)
 XX
 DE Synthetic oligopeptide 5-16' SW.
 XX
 XX Beta-secretase; beta-amyloid protein precursor; APP; Down's syndrome;
 KW Alzheimer's disease; measure activity; cleavage site.
 XX
 OS Synthetic.
 XX
 XX US5942400-A.
 PN
 XX 24-AUG-1999.
 PD
 XX
 XX 07-JUN-1996; 96US-0659984.
 PF
 XX
 XX 07-JUN-1996; 96US-0659984.
 PR
 XX 07-JUN-1995; 95US-0480498.
 PR
 XX 07-JUN-1995; 95US-0485152.
 PR
 XX (ELAN-) ELAN PHARM INC.
 PA
 XX
 XX Anderson JP, Jacobson-Croak KL, Sinha S;
 PI
 XX WPI; 1999-517417/43.
 DR
 XX
 XX A method for detecting human beta-secretase cleavage of polypeptides
 PT useful for identifying beta-secretase inhibitors
 XX
 XX Examples; Column 30; 43pp; English.
 PS
 XX Sequences AAY33752-Y33756 are synthetic oligopeptides used for measuring
 CC the activity of beta-secretase (AAY33741). Beta-secretase is capable of
 CC cleaving beta-amyloid protein precursor (APP) (AAY33742). These
 CC synthetic peptides contain the cleavage site of APP. Beta-secretase and
 CC APP are used in a method for detecting human beta-secretase cleavage of

CC polypeptides and for identifying beta-secretase inhibitors. Inhibition
 CC of beta-secretase activity would be useful for chemical modelling of a
 CC critical event in the pathology of Alzheimer's disease. Inhibitors of
 CC beta-secretase would be useful for the prevention and treatment of
 CC Alzheimer's disease and Down's Syndrome.
 XX
 SQ Sequence 21 AA;
 Query Match 100.0%; Score 49; DB 20; Length 21;
 Best Local Similarity 100.0%; Pred. No. 0.0013;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SEVNLDAEFR 10
 DB 1 sevnlदैfr 10
 |||||
 RESULT 10
 AAW04402
 ID AAW04402 standard; Protein; 32 AA.
 XX
 AC AAW04402;
 XX
 DT 21-JUL-1997 (first entry)
 XX
 DE Mouse amyloid precursor protein exon 16 mutant ST59.
 XX
 XX Exon 16; murine; mouse; amyloid; precursor; protein; APP;
 KW humanisation; homozygous; heterozygous; human; Abeta; Swedish;
 KW familial; Alzheimer's; disease; FAD; mutation; tool; model;
 KW elucidation; pathology; symptomatology; screen; inhibition;
 KW transgenic; mutant; ST59.
 XX
 OS Mus spp.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 13 /note= "corresponding codon GAA"
 FT Misc-difference 15 /note= "wild type Lys substituted with Asn"
 FT Misc-difference 16 /note= "wild type Met substituted with Leu"
 FT Misc-difference 21 /note= "wild type Gly substituted with Arg"
 FT
 XX W09634097-A1.
 PN
 XX 31-OCT-1996.
 PD
 XX
 XX 26-APR-1996; 96WO-US05824.
 PF
 XX
 XX 23-APR-1996; 96US-0636876.
 PR
 XX 26-APR-1995; 95US-0429207.
 PR
 XX (CEPH-) CEPHALON INC.
 PA
 XX
 XX Hoffman EK, Reaume AG, Scott RW, Siman R, Trusko SP;
 PI
 XX WPI; 1996-497629/49.
 DR
 XX N-PSDB; AAT38667.
 DR
 XX Transgenic mice with humanised amyloid precursor protein gene -
 PT having at least 1-Swedish FAD mutation, useful as tools or models to
 PT elucidate role of human A-beta in Alzheimer's disease
 XX
 XX Example 2; Fig 12; 123pp; English.
 PS
 XX The present sequence is the protein encoded by a partial sequence
 CC from exon 16 of the mouse amyloid precursor protein (APP) gene,
 CC into which 4 humanising base pair changes have been introduced. The
 CC exon was then used in the preparation of mice homozygous or
 CC heterozygous for a targeted APP encoding gene, comprising a human

CC	will permit chemical modelling of a critical event in the pathology of	
CC	Alzheimer's disease.	
XX		
SQ	Sequence 33 AA;	
	Query Match 100.0%; Score 49; DB 18; Length 33;	
	Best Local Similarity 100.0%; Pred. No. 0.0023;	
	Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 SEVNLDAEFR 10	
Db	13 sevnldaefr 22	
RESULT 12		
AAV33753		
ID	AAV33753 standard; Protein; 33 AA.	
XX		
AC	AAV33753;	
XX		
DT	09-NOV-1999 (first entry)	
XX		
DE	Synthetic oligopeptide 17-16'SW.	
XX		
XX	Beta-secretase; beta-amyloid protein precursor; APP; Down's syndrome;	
KW	Alzheimer's disease; measure activity; cleavage site.	
OS	Synthetic.	
XX		
PN	US5942400-A.	
XX		
PD	24-AUG-1999.	
XX		
PF	07-JUN-1996; 96US-0659984.	
XX		
PR	07-JUN-1996; 96US-0659984.	
PR	07-JUN-1995; 95US-0480498.	
PR	07-JUN-1995; 95US-0485152.	
XX		
PA	(ELAN-) ELAN PHARM INC.	
XX		
PI	Anderson JP, Jacobson-Croak KL, Sinha S;	
XX		
DR	WPI; 1999-517417/43.	
XX		
PT	A method for detecting human beta-secretase cleavage of polypeptides	
PT	useful for identifying beta-secretase inhibitors	
XX		
PS	Examples; Column 30; 43pp; English.	
XX		
CC	Sequences AAV33752-Y33756 are synthetic oligopeptides used for measuring	
CC	the activity of beta-secretase (AAV33741). Beta-secretase is capable of	
CC	cleaving beta-amyloid protein precursor (APP) (AAV33742). These	
CC	synthetic peptides contain the cleavage site of APP. Beta-secretase and	
CC	APP are used in a method for detecting human beta-secretase cleavage of	
CC	polypeptides and for identifying beta-secretase inhibitors. Inhibition	
CC	of beta-secretase activity would be useful for chemical modelling of a	
CC	critical event in the pathology of Alzheimer's disease. Inhibitors of	
CC	beta-secretase would be useful for the prevention and treatment of	
CC	Alzheimer's disease and Down's Syndrome.	

```

RESULT 13
AAY69718
ID AAY69718 standard; peptide; 39 AA.
XX
XX AC AAY69718;
XX
DT 11-APR-2000 (first entry)
XX
DE Beta-APP alpha-secretase substrate [NL]-APP(-20,+20).
XX
KW Nootropic; neuroprotective; beta-amyloid precursor protein; metabolism;
KW cleavage site; beta-secretase; neurodegenerative disease;
KW Alzheimer's disease.
XX
XX Homo sapiens.
OS Synthetic.
OS
PN WO9964587-A1.
XX
PD 16-DEC-1999.
XX
PF 04-JUN-1999; 99WO-FR01326.
XX
PR 05-JUN-1998; 98FR-0007068.
PR 31-MAR-1999; 99US-0122599.
XX
PA (RHON ) RHONE-POULENC RORER SA.
PA (UYPA-) UNIV CURIE PARIS VI P & M.
XX
PI Rholam M, Munoz-Gimenez N, Moutaouakil M, Cohen P, Bertrand P;
PI
DR WPI; 2000-097537/08.
XX
PT Polypeptide with beta-secretase activity, specific for wild-type
PT amyloid precursor protein, useful in treating Alzheimer's disease
XX
XX Example 3; Page 24; 44pp; French.
XX
CC Peptides AAY69702-Y69718 represent synthetic peptide substrates for a
CC novel polypeptide with beta-secretase activity that can cleave
CC specifically the natural beta-amyloid precursor protein (bAPP). Normal
CC cleavage of the protein occurs between amino acids Met596-Asp597 and
CC Val636-Ile637 (positions 4-5 and 44-45 of AAY69701). The novel
CC polypeptide is used to identify agents that interact specifically with
CC it. These agents regulate metabolism of APP, particularly they slow down
CC or reduce production of beta-amyloid, so can be used to treat
CC neurodegenerative diseases, particularly Alzheimer's disease.
XX
SQ Sequence 39 AA;

Query Match 100.0%; Score 49; DB 21; Length 39;
Best Local Similarity 100.0%; Pred. No. 0.0028;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVNLDAEFR 10
Db 15 sevnldaefr 24
|||||

RESULT 14
AAW08350
ID AAW08350 standard; peptide; 42 AA.
XX
XX AC AAW08350;
XX
DT 05-SEP-1997 (first entry)
XX
DE Wild type APP beta-cleavage site #2.
XX
KW Beta-cleavage site; beta amyloid precursor protein; APP; beta-secretase;
KW alpha-secretase; proteolytic cleavage; inhibitor; Alzheimer's disease.
XX

XX Homo sapiens.
XX WO9640885-A2.
XX 19-DEC-1996.
XX 07-JUN-1996; 96WO-US09985.
XX 07-JUN-1995; 95US-0485152.
XX 07-JUN-1995; 95US-0480498.
XX (ATHE-) ATHENA NEUROSCIENCES INC.
XX
XX Anderson JP, Chrysler SMS, Jacobson-croak KL, Keim PS;
XX Mcconlogue LC, Sinha S, Tan H;
XX WPI; 1997-052304/05.
XX
XX Beta-secretase which specifically cleaves beta-amyloid precursor
XX protein - useful to screen for inhibitors useful in treatment of
XX Alzheimer's disease
XX
XX Claim 5; Page 61; 92pp; English.
XX
XX AAW08216, AAW08217 and AAW08350 represent beta-cleavage sites from
XX beta-amyloid precursor proteins (APP). These sequences are recognised by
XX the enzyme of the invention. The enzyme of the invention is
XX beta-secretase, and specifically cleaves beta-APP at one of these sites.
XX Normal processing of beta-APP is thought to occur via cleavage between
XX residues 16 and 17 of the beta-amyloid peptide region by an
XX alpha-secretase. Pathogenic processing is thought to occur by
XX beta-secretase cleavage of beta-APP. Beta-secretase activity can be
XX detected and measured using a method of the invention, which detects at
XX least one of the beta-secretase cleavage products formed on cleavage. The
XX method can be used to determine whether a test substance inhibits
XX proteolytic cleavage, by beta-secretase, of beta-APP. Compounds effective
XX to at least partially inhibit beta-secretase activity can be used to
XX inhibit cleavage of beta-APP in cells or mammalian hosts. Isolation and
XX purification of beta-secretase will permit chemical modelling of a
XX critical event in the pathology of Alzheimer's disease.
XX
XX Sequence 42 AA;

Query Match 100.0%; Score 49; DB 18; Length 42;
Best Local Similarity 100.0%; Pred. No. 0.003;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVNLDAEFR 10
Db 22 sevnldaefr 31
|||||

RESULT 15
AAY33752
ID AAY33752 standard; Protein; 42 AA.
XX
XX AC AAY33752;
XX
DT 09-NOV-1999 (first entry)
XX
DE Synthetic oligopeptide 26-26'SW.
XX
KW Beta-secretase; beta-amyloid protein precursor; APP; Down's syndrome;
KW Alzheimer's disease; measure activity; cleavage site.
XX
XX Synthetic.
XX
XX OS
XX US5942400-A.
XX
XX 24-AUG-1999.
XX

```

PF 07-JUN-1996; 96US-0659984.
XX
PR 07-JUN-1996; 96US-0659984.
PR 07-JUN-1995; 95US-0480498.
PR 07-JUN-1995; 95US-0485152.
XX
PA (ELAN-) ELAN PHARM INC.
XX
PI Anderson JP, Jacobson-Croak KL, Sinha S;
XX
XX WPI; 1999-517417/43.
DR
XX
XX
PT A method for detecting human beta-secretase cleavage of polypeptides
PT useful for identifying beta-secretase inhibitors
XX
XX
PS Examples; Column 30; 43pp; English.
XX
XX Sequences AAY33752-Y33756 are synthetic oligopeptides used for measuring
CC the activity of beta-secretase (AAY33741). Beta-secretase is capable of
CC cleaving beta-amyloid protein precursor (APP) (AAY33742). These
CC synthetic peptides contain the cleavage site of APP. Beta-secretase and
CC APP are used in a method for detecting human beta-secretase cleavage of
CC polypeptides and for identifying beta-secretase inhibitors. Inhibition
CC of beta-secretase activity would be useful for chemical modelling of a
CC critical event in the pathology of Alzheimer's disease. Inhibitors of
CC beta-secretase would be useful for the prevention and treatment of
CC Alzheimer's disease and Down's Syndrome.
XX
SQ Sequence 42 AA;

Query Match 100.0%; Score 49; DB 20; Length 42;
Best Local Similarity 100.0%; Pred. No. 0.003;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVNLDAEFR 10
Db 22 sevnldaefr 31
|||||

Search completed: September 6, 2001, 16:43:27
Job time: 356 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:39:35 ; Search time 113.12 Seconds
(without alignments)
1.820 Million cell updates/sec

Title: US-09-603-713-5
Perfect score: 49
Sequence: 1 SEVNLDAEFR 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

- 1: /cgn2_6ptodata/2/1aa/5A_COMB.pep.*
- 2: /cgn2_6ptodata/2/1aa/5B_COMB.pep.*
- 3: /cgn2_6ptodata/2/1aa/6A_COMB.pep.*
- 4: /cgn2_6ptodata/2/1aa/6B_COMB.pep.*
- 5: /cgn2_6ptodata/2/1aa/PCTUS_COMB.pep.*
- 6: /cgn2_6ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	49	100.0	10	2	US-08-659-984A-19
2	49	100.0	10	4	US-08-660-531-19
3	49	100.0	11	5	PCT-US94-07043A-3
4	49	100.0	21	2	US-08-659-984A-18
5	49	100.0	21	4	US-08-660-531-18
6	49	100.0	33	2	US-08-659-984A-16
7	49	100.0	33	4	US-08-660-531-16
8	49	100.0	42	2	US-08-659-984A-15
9	49	100.0	42	4	US-08-660-531-15
10	49	100.0	506	2	US-08-659-984A-21
11	49	100.0	506	4	US-08-660-531-21
12	45	91.8	103	3	US-08-339-708A-12
13	44	89.8	9	3	US-08-802-981-219
14	44	89.8	21	3	US-08-802-981-112
15	44	89.8	27	1	US-08-141-324-12
16	44	89.8	27	1	US-08-541-902-12
17	44	89.8	30	2	US-08-659-984A-17
18	44	89.8	30	4	US-08-660-531-17
19	41	83.7	11	5	PCT-US94-07043A-7
20	41	83.7	12	5	PCT-US94-07043A-2
21	41	83.7	16	5	PCT-US94-07043A-1
22	41	83.7	27	1	US-08-141-324-11
23	41	83.7	27	1	US-08-541-902-11
24	41	83.7	45	1	US-08-462-859A-5
25	41	83.7	45	1	US-08-123-659A-5
26	41	83.7	45	1	US-08-464-247A-5
27	41	83.7	45	1	US-08-464-248A-5

28	41	83.7	58	1	US-08-371-930-25	Sequence 25, Appl
29	41	83.7	58	5	PCT-US94-01712-25	Sequence 25, Appl
30	41	83.7	63	1	US-08-462-859A-3	Sequence 3, Appl
31	41	83.7	63	1	US-08-462-859A-4	Sequence 4, Appl
32	41	83.7	63	1	US-08-123-659A-3	Sequence 3, Appl
33	41	83.7	63	1	US-08-123-659A-4	Sequence 4, Appl
34	41	83.7	63	1	US-08-464-247A-3	Sequence 3, Appl
35	41	83.7	63	1	US-08-464-247A-4	Sequence 4, Appl
36	41	83.7	63	1	US-08-464-248A-3	Sequence 3, Appl
37	41	83.7	63	1	US-08-464-248A-4	Sequence 4, Appl
38	41	83.7	105	2	US-08-729-345-1	Sequence 1, Appl
39	41	83.7	117	2	US-08-729-345-3	Sequence 3, Appl
40	41	83.7	152	6	5187153-4	Patent No. 5187153
41	41	83.7	162	6	5220013-4	Patent No. 5220013
42	41	83.7	162	6	5223482-4	Patent No. 5223482
43	41	83.7	264	1	US-07-990-893-5	Sequence 5, Appl
44	41	83.7	487	1	US-08-462-859A-9	Sequence 9, Appl
45	41	83.7	487	1	US-08-123-659A-9	Sequence 9, Appl

ALIGNMENTS

RESULT 1
US-08-659-984A-19
; Sequence 19, Application US/08659984A
; Patent No. 5942400
; GENERAL INFORMATION:
; APPLICANT: Anderson, John P.
; APPLICANT: Sinha, Sukanto
; APPLICANT: Jacobson-Croak, Kirsten L.
; TITLE OF INVENTION: Assays for Detecting Beta-Secretase
; TITLE OF INVENTION: Inhibition
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Ctr., 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/659,984A
; FILING DATE: 07-JUN-1996
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,152
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Heslin, James M.
; REGISTRATION NUMBER: 29,541
; REFERENCE/DOCKET NUMBER: 15270-002810US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Region
; LOCATION: one-of(1)
; OTHER INFORMATION: /note= "N-terminal Ser is acetylated."
US-08-659-984A-19

Query Match 100.0%; Score 49; DB 2; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.00029;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SEVNLDAEFR 10
Db 1 SEVNLDAEFR 10

RESULT 2
US-08-660-531-19
; Sequence 19, Application US/08660531
; Patent No. 6221845
; GENERAL INFORMATION:
; APPLICANT: Chrysler, Susanna M.S.
; APPLICANT: Sinha, Sukanto
; APPLICANT: Keim, Pamela S.
; APPLICANT: Anderson, John P.
; TITLE OF INVENTION: Beta-Secretase
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Crew LLP
; STREET: Two Embarcadero Ctr., 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/660.531
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/480.498
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Heslin, James M.
; REGISTRATION NUMBER: 29,541
; REFERENCE/DOCKET NUMBER: 15270-002210US
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Region
; LOCATION: one-of(1)
; OTHER INFORMATION: /note= "N-terminal Ser is acetylated."
; US-08-660-531-19

Query Match 100.0%; Score 49; DB 4; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.00029;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SEVNLDAEFR 10
Db 1 SEVNLDAEFR 10

RESULT 3
PCT-US94-07043A-3
; Sequence 3, Application PC/TUS9407043A

Query Match 100.0%; Score 49; DB 5; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.00033;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SEVNLDAEFR 10
Db 2 SEVNLDAEFR 11

RESULT 4
US-08-659-984A-18
; Sequence 18, Application US/08659984A
; Patent No. 5942400
; GENERAL INFORMATION:
; APPLICANT: Anderson, John P.
; APPLICANT: Sinha, Sukanto
; APPLICANT: Jacobson-Croak, Kirsten L.
; TITLE OF INVENTION: Assays for Detecting Beta-Secretase
; TITLE OF INVENTION: Inhibition
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Crew LLP
; STREET: Two Embarcadero Ctr., 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834

GENERAL INFORMATION:
; APPLICANT: Tamburini, Paul P.; Benz, G nter; H bich,
; APPLICANT: Dieter; Drever, Robert N.; Koenig, Gerhard
; TITLE OF INVENTION: CATHEPSIN D IS AN AMYLOIDGENIC
; TITLE OF INVENTION: PROTEASE IN ALZHEIMER S DISEASE
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Miles Inc.
; STREET: 400 Morgan Lane
; CITY: West Haven
; STATE: Connecticut
; COUNTRY: USA
; ZIP: 06516
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 800 kb storage
; COMPUTER: Sharp PC 4600
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/07043A
; FILING DATE: June 21, 1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/10889
; FILING DATE: November 12, 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/995,660
; FILING DATE: December 16, 1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/880,914
; FILING DATE: May 11, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Pamela A. Simonton
; REGISTRATION NUMBER: 31,060
; REFERENCE/DOCKET NUMBER: MTI 224.3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (203) 937-2340
; TELEFAX: (203) 937-2795
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; PCT-US94-07043A-3

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/659,984A
FILING DATE: 07-JUN-1996
CLASSIFICATION: 436
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/485,152
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Heslin, James M.
REGISTRATION NUMBER: 29,541
REFERENCE/DOCKET NUMBER: 15270-002810US
TELEPHONE: 415-326-2400
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-659-984A-18

Query Match 100.0%; Score 49; DB 2; Length 21;
Best Local Similarity 100.0%; Pred. No. 0.00069;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SEVNLDAEFR 10
Db 1 SEVNLDAEFR 10

RESULT 5
US-08-660-531-18
Sequence 18, Application US/08660531
Patent No. 6221645
GENERAL INFORMATION:
APPLICANT: Chrysler, Susanna M.S.
APPLICANT: Sinha, Sukanto
APPLICANT: Keim, Pamela S.
APPLICANT: Anderson, John P.
TITLE OF INVENTION: Beta-Secretase
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Ctr., 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/660,531
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/480,498
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Heslin, James M.
REGISTRATION NUMBER: 29,541
REFERENCE/DOCKET NUMBER: 15270-002210US
TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-326-2400
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-660-531-18

Query Match 100.0%; Score 49; DB 4; Length 21;
Best Local Similarity 100.0%; Pred. No. 0.00069;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SEVNLDAEFR 10
Db 1 SEVNLDAEFR 10

RESULT 6
US-08-659-984A-16
Sequence 16, Application US/08659984A
Patent No. 5942400
GENERAL INFORMATION:
APPLICANT: Anderson, John P.
APPLICANT: Sinha, Sukanto
APPLICANT: Jacobson-Croak, Kirsten L.
TITLE OF INVENTION: Assays for Detecting Beta-Secretase
TITLE OF INVENTION: Inhibition
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Ctr., 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/659,984A
FILING DATE: 07-JUN-1996
CLASSIFICATION: 436
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/485,152
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Heslin, James M.
REGISTRATION NUMBER: 29,541
REFERENCE/DOCKET NUMBER: 15270-002810US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-326-2400
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 33 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-659-984A-15

Query Match 100.0%; Score 49; DB 2; Length 33;
Best Local Similarity 100.0%; Pred. No. 0.0011;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SEVNLDAEFR 10

Db 13 SEVNLDAEFR 22

RESULT 7
US-08-660-531-16
; Sequence 16, Application US/08660531
; Patent No. 6221645
; GENERAL INFORMATION:
; APPLICANT: Chrysler, Susanna M.S.
; APPLICANT: Sinha, Sukanto
; APPLICANT: Keim, Pamela S.
; APPLICANT: Anderson, John P.
; TITLE OF INVENTION: Beta-Secretase
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Ctr., 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/660,531
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/480,498
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Heslin, James M.
; REGISTRATION NUMBER: 29,541
; REFERENCE/DOCKET NUMBER: 15270-002210US
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 33 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-660-531-16

Query Match 100.0%; Score 49; DB 4; Length 33;
Best Local Similarity 100.0%; Pred. No. 0.0011;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVNLDAEFR 10
Db 13 SEVNLDAEFR 22

RESULT 8
US-08-659-984A-15
; Sequence 15, Application US/08659984A
; Patent No. 5942400
; GENERAL INFORMATION:
; APPLICANT: Anderson, John P.
; APPLICANT: Sinha, Sukanto
; APPLICANT: Jacobson-Croak, Kirsten L.
; TITLE OF INVENTION: Assays for Detecting Beta-Secretase
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP

; STREET: Two Embarcadero Ctr., 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/659,984A
; FILING DATE: 07-JUN-1996
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,152
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Heslin, James M.
; REGISTRATION NUMBER: 29,541
; REFERENCE/DOCKET NUMBER: 15270-002810US
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-659-984A-15

Query Match 100.0%; Score 49; DB 2; Length 42;
Best Local Similarity 100.0%; Pred. No. 0.0015;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVNLDAEFR 10
Db 22 SEVNLDAEFR 31

RESULT 9
US-08-660-531-15
; Sequence 15, Application US/08660531
; Patent No. 6221645
; GENERAL INFORMATION:
; APPLICANT: Chrysler, Susanna M.S.
; APPLICANT: Sinha, Sukanto
; APPLICANT: Keim, Pamela S.
; APPLICANT: Anderson, John P.
; TITLE OF INVENTION: Beta-Secretase
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Ctr., 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/660,531
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/480,498
; FILING DATE: 07-JUN-1995

ATTORNEY/AGENT INFORMATION:
NAME: Heslin, James M.
REGISTRATION NUMBER: 29,541
REFERENCE/DOCKET NUMBER: 15270-002210US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-326-2400
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 42 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-660-531-15

Query Match 100.0%; Score 49; DB 4; Length 42;
Best Local Similarity 100.0%; Pred. No. 0.0015;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVNLDAEFR 10
DB 22 SEVNLDAEFR 31

RESULT 10
US-08-659-984A-21
; Sequence 21, Application US/08659984A
; Patent No. 5942400
; GENERAL INFORMATION:
; APPLICANT: Anderson, John P.
; APPLICANT: Sinha, Sukanto
; APPLICANT: Jacobson-Croak, Kirsten L.
; TITLE OF INVENTION: Assays for Detecting Beta-Secretase
; TITLE OF INVENTION: Inhibition
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Ctr., 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/659,984A
; FILING DATE: 07-JUN-1996
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,152
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Heslin, James M.
; REGISTRATION NUMBER: 29,541
; REFERENCE/DOCKET NUMBER: 15270-002810US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 506 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-659-984A-21

Query Match 100.0%; Score 49; DB 2; Length 506;

Best Local Similarity 100.0%; Pred. No. 0.025;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVNLDAEFR 10
DB 403 SEVNLDAEFR 412

RESULT 11
US-08-660-531-21
; Sequence 21, Application US/08660531
; Patent No. 6221645
; GENERAL INFORMATION:
; APPLICANT: Chrysler, Susanna M.S.
; APPLICANT: Sinha, Sukanto
; APPLICANT: Keim, Pamela S.
; APPLICANT: Anderson, John P.
; TITLE OF INVENTION: Beta-Secretase
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Ctr., 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/660,531
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/480,498
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Heslin, James M.
; REGISTRATION NUMBER: 29,541
; REFERENCE/DOCKET NUMBER: 15270-002210US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 506 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-660-531-21

Query Match 100.0%; Score 49; DB 4; Length 506;
Best Local Similarity 100.0%; Pred. No. 0.025;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVNLDAEFR 10
DB 403 SEVNLDAEFR 412

RESULT 12
US-08-339-708A-12
; Sequence 12, Application US/08339708A
; Patent No. 6037521
; GENERAL INFORMATION:
; APPLICANT: Sato, Masahiro
; APPLICANT: Takashi, Kobayashi
; APPLICANT: Tada, No. 6037521hiro
; APPLICANT: Shoji, Mikio
; APPLICANT: Kawarabayashi, Takeshi

```

; TITLE OF INVENTION: TRANSGENIC ANIMAL MODEL FOR ALZHEIMER'S
; TITLE OF INVENTION: DISEASE
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/339,708A
; FILING DATE: 14-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 306026/93
; FILING DATE: 12-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: SANDERCOCK, COLIN G.
; REGISTRATION NUMBER: 31,298
; REFERENCE/DOCKET NUMBER: 026083/0159
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 103 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-339-708A-12

```

```

Query Match          91.8%; Score 45; DB 3; Length 103;
Best Local Similarity 100.0%; Pred. No. 0.027;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2 EVNLDAAEF 10
    |||||
Db 1 EVNLDAAEF 9

```

```

RESULT 13
US-08-802-981-219
; Sequence 219, Application US/08802981
; Patent No. 6037137
; GENERAL INFORMATION:
; APPLICANT: Komoriya, Akira
; APPLICANT: Packard, Beverly S.
; TITLE OF INVENTION: Compositions for the Detection of Enzyme
; TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof
; NUMBER OF SEQUENCES: 231
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/802,981
; FILING DATE: 20-FEB-1997
; CLASSIFICATION: 435

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Hunter, Tom
; REGISTRATION NUMBER: 38,498
; REFERENCE/DOCKET NUMBER: 016865-0003000US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 219:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-802-981-219

```

```

Query Match          89.8%; Score 44; DB 3; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.4e-05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 SEVNLDAEF 9
    |||||
Db 1 SEVNLDAEF 9

```

```

RESULT 14
US-08-802-981-112
; Sequence 112, Application US/08802981
; Patent No. 6037137
; GENERAL INFORMATION:
; APPLICANT: Komoriya, Akira
; APPLICANT: Packard, Beverly S.
; TITLE OF INVENTION: Compositions for the Detection of Enzyme
; TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof
; NUMBER OF SEQUENCES: 231
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/802,981
; FILING DATE: 20-FEB-1997
; CLASSIFICATION: 435

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Hunter, Tom
; REGISTRATION NUMBER: 38,498
; REFERENCE/DOCKET NUMBER: 016865-0003000US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 112:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 3
; OTHER INFORMATION: /product= "Alb"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 4

```

;; OTHER INFORMATION: /product= "Acp"
;; FEATURE:
;; NAME/KEY: Modified-site
;; LOCATION: 16
;; OTHER INFORMATION: /product= "Acp"
US-08-802-981-112

Search completed: September 6, 2001, 16:39:35
Job time: 124 sec

Query Match 89.8%; Score 44; DB 3; Length 21;
Best Local Similarity 100.0%; Pred. No. 0.0071;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 SEVNLDAEF 9
Db 6 SEVNLDAEF 14

RESULT 15
US-08-141-324-12
; Sequence 12, Application US/08141324
; Patent No. 5475097
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S.
; APPLICANT: Barr, Philip J.
; APPLICANT: Pavloff, Nadine
; APPLICANT: Pike, Robert N.
; TITLE OF INVENTION: Lysine-specific Porphyromonas gingivalis
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee and Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: CO
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/141.324
; FILING DATE: 21-OCT-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 44-93
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303-499-8080
; TELEFAX: 303-499-8089
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 27 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-141-324-12

Query Match 89.8%; Score 44; DB 1; Length 27;
Best Local Similarity 90.0%; Pred. No. 0.0094;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 SEVNLDAEFR 10
Db 4 SEVLDAEFR 13

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:45:47 ; Search time 134.15 Seconds
(without alignments)
5.678 Million cell updates/sec

Title: US-09-603-713-5

Perfect score: 49

Sequence: 1 SEVNLDAEFR 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_68.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	41	83.7	57	2	E60045 Alzheimer's diseases
2	41	83.7	57	2	E60045 Alzheimer's diseases
3	41	83.7	57	2	G60045 Alzheimer's diseases
4	41	83.7	57	2	D60045 Alzheimer's diseases
5	41	83.7	57	2	A60045 Alzheimer's diseases
6	41	83.7	57	2	B60045 Alzheimer's diseases
7	41	83.7	82	2	PQ0438 Alzheimer's diseases
8	41	83.7	695	1	A49795 Alzheimer's diseases
9	41	83.7	770	1	ORHUA4 Alzheimer's diseases
10	38	77.6	419	2	D42725 nitrile hydratase
11	38	77.6	4639	1	A54794 dynein heavy chain
12	37	75.5	470	2	C75591 threonine synthase
13	36	73.5	33	2	S23094 beta-amyloid prote
14	36	73.5	695	2	A27485 Alzheimer's diseases
15	36	73.5	695	2	S00550 Alzheimer's diseases
16	35	71.4	265	2	S64127 hypothetical prote
17	35	71.4	282	2	T26112 hypothetical prote
18	35	71.4	434	2	J72430 transcription elon
19	35	71.4	747	2	JH0773 Alzheimer's diseases
20	35	71.4	929	2	T52517 hypothetical prote
21	34	69.4	1024	2	C64208 hypothetical prote
22	33	67.3	52	2	T47295 hypothetical prote
23	33	67.3	279	2	T41124 single-stranded DN
24	33	67.3	345	2	E70162 DNA-directed RNA p
25	33	67.3	442	2	B82633 nitrile hydratase
26	33	67.3	2514	2	T37320 ataxia telangiecta
27	33	67.3	2619	2	T24588 hypothetical prote
28	32	65.3	103	2	T70909 hypothetical prote
29	32	65.3	197	2	E71692 NADH dehydrogenase

hypothetical prote
MTH1 protein - yea
plasma phospholipi
probable membrane
hypothetical prote
hypothetical prote
glycoprotein H pre
probable chromosom
hypothetical prote
chromosome scaffold
hypothetical prote
laminin alpha-1 ch
hypothetical prote
hypothetical prote
hypothetical prote
probable amidase 1

ALIGNMENTS

RESULT 1

E60045

Alzheimer's disease amyloid beta/A4 protein precursor - sheep (fragment)

C:Species: Ovis sp. (sheep)

C:Date: 01-Dec-1992 #sequence_revision 01-Dec-1992 #text_change 28-Jul-1995

C:Accession: E60045

R:Johnstone, E.M.; Chaney, M.O.; Norris, F.H.; Pascual, R.; Little, S.P.

Brain Res. Mol. Brain Res. 10, 299-305, 1991

A:Title: Conservation of the sequence of the Alzheimer's disease amyloid peptide in d

A:Reference number: A60045; MUID:92017079

A:Accession: E60045

A:Molecule type: mRNA

A:Residues: 1-57 <JOH>

A:Cross-references: EMBL:X56130

C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type protein

C:Keywords: alternative splicing; Alzheimer's disease; amyloid; brain

Query Match 83.7%; Score 41; DB 2; Length 57;

Best Local Similarity 80.0%; Pred. No. 0.087;

Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SEVNLDAEFR 10

Db 1 SEVKMDAEFR 10

RESULT 2

F60045

Alzheimer's disease amyloid beta/A4 protein precursor - pig (fragment)

C:Species: Sus scrofa domestica (domestic pig)

C:Date: 01-Dec-1992 #sequence_revision 01-Dec-1992 #text_change 13-Aug-1999

C:Accession: F60045

R:Johnstone, E.M.; Chaney, M.O.; Norris, F.H.; Pascual, R.; Little, S.P.

Brain Res. Mol. Brain Res. 10, 299-305, 1991

A:Title: Conservation of the sequence of the Alzheimer's disease amyloid peptide in d

A:Reference number: A60045; MUID:92017079

A:Accession: F60045

A:Molecule type: mRNA

A:Residues: 1-57 <JOH>

A:Cross-references: EMBL:X56127; NID:g1895; PIDN:CAA39592.1; PID:g1896

C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type protein

C:Keywords: alternative splicing; Alzheimer's disease; amyloid; brain

Query Match 83.7%; Score 41; DB 2; Length 57;

Best Local Similarity 80.0%; Pred. No. 0.087;

Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SEVNLDAEFR 10

Db 1 SEVKMDAEFR 10

RESULT 3
G60045
Alzheimer's disease amyloid beta/A4 protein precursor - guinea pig (fragment)
C:Species: Cavia porcellus (guinea pig)
C:Date: 01-Dec-1992 #sequence_revision 01-Dec-1992 #text_change 28-Jul-1995
C:Accession: G60045
R:Johnstone, E.M.; Chaney, M.O.; Norris, F.H.; Pascual, R.; Little, S.P.
Brain Res. Mol. Brain Res. 10, 299-305, 1991
A:Title: Conservation of the sequence of the Alzheimer's disease amyloid peptide in dog,
A:Accession: G60045
A:Reference number: A60045; MUID:92017079
A:Molecule type: mRNA
A:Residues: 1-57 <JOH>
A:Cross-references: EMBL:X56126
C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinase
C:Keywords: alternative splicing; Alzheimer's disease; amyloid; brain

Query Match 83.7%; Score 41; DB 2; Length 57;
Best Local Similarity 80.0%; Pred. No. 0.087;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEVNLDAEFR 10
||| :|||||
Db 1 SEVKMDAEFR 10

RESULT 4
D60045
Alzheimer's disease amyloid beta/A4 protein precursor - bovine (fragment)
C:Species: Bos primigenius taurus (cattle)
C:Date: 01-Dec-1992 #sequence_revision 01-Dec-1992 #text_change 28-Jul-1995
C:Accession: D60045
R:Johnstone, E.M.; Chaney, M.O.; Norris, F.H.; Pascual, R.; Little, S.P.
Brain Res. Mol. Brain Res. 10, 299-305, 1991
A:Title: Conservation of the sequence of the Alzheimer's disease amyloid peptide in dog,
A:Reference number: A60045; MUID:92017079
A:Accession: D60045
A:Molecule type: mRNA
A:Residues: 1-57 <JOH>
A:Cross-references: EMBL:X56124
C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinase
C:Keywords: alternative splicing; Alzheimer's disease; amyloid; brain

Query Match 83.7%; Score 41; DB 2; Length 57;
Best Local Similarity 80.0%; Pred. No. 0.087;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEVNLDAEFR 10
||| :|||||
Db 1 SEVKMDAEFR 10

RESULT 5
A60045
Alzheimer's disease amyloid beta/A4 protein precursor - dog (fragment)
C:Species: Canis lupus familiaris (dog)
C:Date: 01-Dec-1992 #sequence_revision 01-Dec-1992 #text_change 28-Jul-1995
C:Accession: A60045
R:Johnstone, E.M.; Chaney, M.O.; Norris, F.H.; Pascual, R.; Little, S.P.
Brain Res. Mol. Brain Res. 10, 299-305, 1991
A:Title: Conservation of the sequence of the Alzheimer's disease amyloid peptide in dog,
A:Reference number: A60045; MUID:92017079
A:Accession: A60045
A:Molecule type: mRNA
A:Residues: 1-57 <JOH>
A:Cross-references: EMBL:X56125
C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinase
C:Keywords: alternative splicing; Alzheimer's disease; amyloid; brain

Query Match 83.7%; Score 41; DB 2; Length 57;
Best Local Similarity 80.0%; Pred. No. 0.087;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEVNLDAEFR 10
||| :|||||
Db 1 SEVKMDAEFR 10

Query Match 83.7%; Score 41; DB 2; Length 57;
Best Local Similarity 80.0%; Pred. No. 0.087;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEVNLDAEFR 10
||| :|||||
Db 1 SEVKMDAEFR 10

RESULT 6
B60045
Alzheimer's disease amyloid beta/A4 protein precursor - polar bear (fragment)
C:Species: Ursus maritimus (polar bear)
C:Date: 01-Dec-1992 #sequence_revision 01-Dec-1992 #text_change 13-Aug-1999
C:Accession: B60045
R:Johnstone, E.M.; Chaney, M.O.; Norris, F.H.; Pascual, R.; Little, S.P.
Brain Res. Mol. Brain Res. 10, 299-305, 1991
A:Title: Conservation of the sequence of the Alzheimer's disease amyloid peptide in d
A:Reference number: A60045; MUID:92017079
A:Accession: B60045
A:Molecule type: mRNA
A:Residues: 1-57 <JOH>
A:Cross-references: EMBL:X56128; NID:92165; PIDN:CAA39593.1; PID:92166
C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinase
C:Keywords: alternative splicing; Alzheimer's disease; amyloid; brain

Query Match 83.7%; Score 41; DB 2; Length 57;
Best Local Similarity 80.0%; Pred. No. 0.087;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEVNLDAEFR 10
||| :|||||
Db 1 SEVKMDAEFR 10

RESULT 7
PQ0438
Alzheimer's disease amyloid A4 protein precursor - rabbit (fragment)
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 30-Sep-1993 #sequence_revision 19-Oct-1995 #text_change 19-Oct-1995
C:Accession: PQ0438; C60045
R:Davidson, J.S.; West, R.L.; Kotikalapudi, P.; Maroun, L.E.
Biochem. Biophys. Res. Commun. 189, 905-911, 1992
A:Title: Sequence and methylation in the beta/A4 region of the rabbit amyloid precurs
A:Reference number: PQ0438; MUID:93075180
A:Accession: PQ0438
A:Molecule type: DNA
A:Residues: 1-82 <DAV>
A:Cross-references: GB:M83558; GB:M83657
R:Johnstone, E.M.; Chaney, M.O.; Norris, F.H.; Pascual, R.; Little, S.P.
Brain Res. Mol. Brain Res. 10, 299-305, 1991
A:Title: Conservation of the sequence of the Alzheimer's disease amyloid peptide in d
A:Reference number: A60045; MUID:92017079
A:Accession: C60045
A:Molecule type: mRNA
A:Residues: 12-68 <JOH>
A:Cross-references: EMBL:X56129
C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinase
C:Keywords: alternative splicing; Alzheimer's disease; amyloid; Down's syndrome

Query Match 83.7%; Score 41; DB 2; Length 82;
Best Local Similarity 80.0%; Pred. No. 0.13;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEVNLDAEFR 10
||| :|||||
Db 12 SEVKMDAEFR 21

RESULT 8

A49795
 Alzheimer's disease amyloid beta protein precursor - crab-eating macaque
 C:Species: Macaca fascicularis (crab-eating macaque)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: A49795
 R:Podlasky, M.B.; Tolan, D.R.; Selkoe, D.J.
 Am. J. Pathol. 136, 1423-1435, 1991
 A:Title: Homology of the amyloid beta protein precursor in monkey and human supports a beta-amyloid precursor hypothesis
 A:Reference number: A49795; MUID:91273117
 A:Accession: A49795
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-695 <POD>
 A:Cross-references: GB:M58727; NID:g342062; PIDN:AAA36829.1; PID:g342063
 C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinase 1
 C:Keywords: alternative splicing

Query Match 83.7%; Score 41; DB 1; Length 695;
 Best Local Similarity 80.0%; Pred. NO. 1.6;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 SEVNLDAEER 10
 ||| :|||
 Db 592 SEVKMDAEER 601

RESULT 9
 ORHU44
 Alzheimer's disease amyloid beta protein precursor [validated] - human
 N:Alternate names: Alzheimer's disease amyloid A4 protein; coagulation factor XIA inhibitor
 N:Contents: amyloid beta protein long, plaque form; amyloid beta protein short, vascular
 protein precursor splice form APP(770)
 C:Species: Homo sapiens (man)
 C:Date: 30-Jun-1987 #sequence_revision 28-Jul-1995 #text_change 15-Sep-2000
 C:Accession: S02260; S05194; A32277; A33260; A3486; I39452; I39451; I39453; I59562; A44
 4668; A28583; A29302; A60805; J00338; S06121; A60355; A59011; A38384; S29076; S38252; S3
 R:Lemaire, H.G.; Salbaum, J.M.; Multhaup, G.; Kang, J.; Bayney, R.M.; Unterbeck, A.; Bey
 Nucleic Acids Res. 17, 517-522, 1989
 A:Title: The PreA4(695) precursor protein of Alzheimer's disease A4 amyloid is encoded b
 A:Reference number: S02260; MUID:89128427
 A:Accession: S02260
 A:Molecule type: DNA
 A:Residues: 1-288, 'V', 365-770 <LEM1>
 A:Cross-references: EMBL:X13466
 A:Note: alternative splice form APP(695)
 R:Lemaire, H.G.
 submitted to the EMBL Data Library, November 1988
 A:Reference number: S05194
 A:Accession: S05194
 A:Molecule type: DNA
 A:Residues: 1-14, 'VW', 17-288, 'V', 365-770 <LEM2>
 A:Cross-references: EMBL:X13466; NID:g35598; PIDN:CAA31830.1; PID:g871360
 A:Note: alternative splice form APP(695)
 R:La Faurie, G.; Lahiri, D.K.; Salton, S.R.J.; Robakis, N.K.
 Biochem. Biophys. Res. Commun. 159, 297-304, 1989
 A:Title: Characterization of the 5'-end region and the first two exons of the beta-prote
 A:Reference number: A32277; MUID:89165870
 A:Accession: A32277
 A:Molecule type: DNA
 A:Residues: 1-75 <LAF>
 A:Cross-references: GB:M24546; GB:M24547; NID:g341202; PIDN:AAC13654.1; PID:g516074
 R:Johnstone, E.M.; Chaney, M.O.; Moore, R.E.; Ward, K.E.; Norris, F.H.; Little, S.P.
 Biochem. Biophys. Res. Commun. 163, 1248-1255, 1989
 A:Title: Alzheimer's disease amyloid peptide is encoded by two exons and shows similarit
 A:Reference number: A33260; MUID:89392030
 A:Accession: A33260
 A:Molecule type: DNA
 A:Residues: 656-737 <JOH>
 A:Cross-references: GB:M29270; NID:g178863; PIDN:AAA51768.1; PID:g178865
 R:Prelli, F.; Levy, E.; van Duinen, S.G.; Bots, G.T.A.M.; Luyendijk, W.; Frangione, B.
 Biochem. Biophys. Res. Commun. 170, 301-307, 1990
 A:Title: Expression of a normal and variant Alzheimer's beta-protein gene in amyloid of

A:Reference number: A35486; MUID:90321244
 A:Accession: A35486
 A:Molecule type: DNA
 A:Residues: 672-710 <PRE1>
 A:Note: 693-Gln was found in DNA isolated from HCHWA-D patients
 R:Yoshikai, S.I.; Sasaki, H.; Doh-ura, K.; Furuya, H.; Sasaki, Y.
 Gene 87, 257-263, 1990
 A:Title: Genomic organization of the human amyloid beta-protein precursor gene.
 A:Reference number: I39451; MUID:90236318
 A:Accession: I39452
 A:Status: nucleic acid sequence not shown; translation not shown; translated from GB/
 A:Molecule type: DNA
 A:Residues: 1-770 <YOS1>
 A:Cross-references: GB:M33112; NID:g178613; PIDN:AAB59502.1; PID:g178616
 A:Accession: I39451
 A:Status: nucleic acid sequence not shown; translation not shown; translated from GB/
 A:Molecule type: DNA
 A:Residues: 1-530, 'QMLMPVIPAFWEAKVGR' <YOS2>
 A:Cross-references: GB:M34875; NID:g178608; PIDN:AAB59501.1; PID:g178615
 R:Yoshikai, S.I.; Sasaki, H.; Doh-ura, K.; Furuya, H.; Sasaki, Y.
 Gene 102, 291-292, 1991
 A:Reference number: A59020; MUID:91340168
 A:Contents: annotation: erratum
 A:Note: revised physical map for reference I39451
 R:Levy, E.; Carman, M.D.; Fernandez-Madrid, I.J.; Power, M.D.; Lieberburg, I.; van Du
 Science 248, 1124-1126, 1990
 A:Title: Mutation of the Alzheimer's disease amyloid gene in hereditary cerebral hemo
 A:Reference number: I39453; MUID:90260663
 A:Accession: I39453
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 656-737 <LEV>
 A:Cross-references: GB:M37896; NID:g178618; PIDN:AAA51727.1; PID:g178620
 R:Murrell, J.; Farlow, M.; Ghetti, B.; Benson, M.D.
 Science 254, 97-99, 1991
 A:Title: A mutation in the amyloid precursor protein associated with hereditary Alzhe
 A:Reference number: I59562; MUID:92022553
 A:Accession: I59562
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 689-716, 'P', 718-737 <MUR>
 A:Cross-references: GB:S57665; NID:g236720; PIDN:AAB19991.1; PID:g236721
 R:Kamino, K.; Orr, H.T.; Payami, H.; Wijsman, E.M.; Alonso, M.E.; Pulst, S.M.; Anders
 arakis, S.E.; Korenberg, J.R.; Sharma, V.; Kukull, W.; Larson, E.; Heston, L.L.; Mart
 Am. J. Hum. Genet. 51, 998-1014, 1992
 A:Title: Linkage and mutational analysis of familial Alzheimer disease kindreds for t
 A:Reference number: A44017; MUID:93035397
 A:Accession: A44017
 A:Molecule type: DNA
 A:Residues: 687-692, 'G', 694-718 <KAM1>
 A:Cross-references: GB:S45135; NID:g257377; PIDN:AAB23645.1; PID:g257378
 A:Experimental source: familial Alzheimer disease family SB
 A:Note: sequence extracted from NCBI backbone (NCBIP:115374)
 A:Accession: B44017
 A:Molecule type: DNA
 A:Residues: 687-718 <KAM2>
 A:Cross-references: GB:S45136; NID:g257379; PIDN:AAB23646.1; PID:g257380
 A:Experimental source: familial Alzheimer disease family LIT
 A:Note: sequence extracted from NCBI backbone (NCBIP:115376)
 A:Note: this sequence has a silent mutation
 R:Kang, J.; Lemaire, H.G.; Unterbeck, A.; Salbaum, J.M.; Masters, C.L.; Grzeschlk, K.
 Nature 325, 733-736, 1987
 A:Title: The precursor of Alzheimer's disease amyloid A4 protein resembles a cell-sur
 A:Reference number: A03134; MUID:87144572
 A:Accession: A03134
 A:Molecule type: mRNA
 A:Residues: 1-288, 'V', 365-770 <KAN>
 A:Cross-references: GB:Y00264; NID:g28525; PIDN:CAA68374.1; PID:g28526
 A:Note: alternative splice form APP(695)
 R:Robakis, N.K.; Ramakrishna, N.; Wolfe, G.; Wisniewski, H.M.
 Proc. Natl. Acad. Sci. U.S.A. 84, 4190-4194, 1987
 A:Title: Molecular cloning and characterization of a cDNA encoding the cerebrovascula

A:Reference number: A29030; MUID:87231971
A:Accession: A29030
A:Molecule type: mRNA
A:Residues: 284-288,'V',365-646,'E',648-770 <ROB>
A:Cross-references: GB:M16765; NID:g178539; PIDN:AAA51722.1; PID:g178540
A:Note: the authors translated the codon GAG for residue 647 as Asp
R:Goldgaber, D.; Lerman, M.I.; McBride, O.W.; Saffioti, U.; Gajdusek, D.C.
Science 235, 877-880, 1987
A:Title: Characterization and chromosomal localization of a cDNA encoding brain amyloid
A:Reference number: A47584; MUID:87120328
A:Accession: A47584
A:Molecule type: mRNA
A:Residues: 674-756,'S',758-770 <GOL>
A:Cross-references: GB:M15533; NID:g178706; PIDN:AAA35540.1; PID:g178707
A:Experimental source: brain
R:Tanzi, R.E.; Gusella, J.F.; Watkins, P.C.; Bruns, G.A.P.; St George-Hyslop, P.; Van Ke
Science 235, 880-884, 1987
A:Title: Amyloid beta protein gene: cDNA, mRNA distribution, and genetic linkage near th
A:Reference number: A47585; MUID:87120329
A:Accession: A47585
A:Molecule type: mRNA
A:Residues: 674-703 <TANL>
A:Cross-references: GB:M15532; NID:g177957; PIDN:AAA51564.1; PID:g177958
R:Dyrks, T.; Weidemann, A.; Multhaup, G.; Salbaum, J.M.; Lemaire, H.G.; Kang, J.; Muelle
EMBO J. 7, 949-957, 1988
A:Title: Identification, transmembrane orientation and biogenesis of the amyloid A4 prec
A:Reference number: S02638; MUID:88296437
A:Accession: S02638
A:Molecule type: mRNA
A:Residues: 672-678 <DYS>
R:Tanzi, R.E.; McClatchey, A.I.; Lamperti, E.D.; Villa-Komaroff, L.; Gusella, J.F.; Neve
Nature 331, 528-530, 1988
A:Title: Protease inhibitor domain encoded by an amyloid protein precursor mRNA associat
A:Reference number: S00707; MUID:88122640
A:Accession: S00707
A:Molecule type: mRNA
A:Residues: 286-344,'I',365-366 <TAN2>
A:Cross-references: EMBL:X06982; NID:g28817; PIDN:CAA30042.1; PID:g2929612
A:Experimental source: promyelocytic leukemia cell line HL60
A:Note: alternative splice form APP(751)
R:Ponte, P.; Gonzalez-Dewhitt, P.; Schilling, J.; Miller, J.; Hsu, D.; Greenberg, B.; Da
Nature 331, 525-527, 1988
A:Title: A new A4 amyloid mRNA contains a domain homologous to serine proteinase inhibit
A:Reference number: S00925; MUID:88122639
A:Accession: S00925
A:Molecule type: mRNA
A:Residues: 1-344,'I',365-770 <PO2>
A:Cross-references: EMBL:X06989; EMBL:Y00297; NID:g28720; PIDN:CAA30050.1; PID:g28721
A:Note: alternative splice form APP(751)
R:Kitaguchi, N.; Takahashi, Y.; Tokushima, Y.; Shiojiri, S.; Ito, H.
Nature 331, 530-532, 1988
A:Title: Novel precursor of Alzheimer's disease amyloid protein shows protease inhibitor
A:Reference number: A38949; MUID:88122641
A:Accession: A38949
A:Molecule type: mRNA
A:Residues: 287-367 <KIT>
A:Cross-references: GB:X06981; NID:g28816; PIDN:CAA30041.1; PID:g929611
A:Experimental source: glioblastoma cell line
A:Note: alternative splice form APP(770)
R:Vitek, M.P.; Rascol, C.G.; de Sauvage, F.; Vitek, S.M.; Bartus, R.T.; Beer, B.; Ashton
Brain Res. Mol. Brain Res. 4, 121-131, 1988
A:Title: Absence of mutation in the beta-amyloid cDNAs cloned from the brains of three P
A:Reference number: A30320
A:Accession: A30320
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 284-288,'V',365-770 <VIT1>
A:Accession: B30320
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 122-288,'V',365-770 <VIT2>
A:Accession: C30320
A>Status: not compared with conceptual translation

A:Molecule type: mRNA
A:Residues: 606-770 <VIT3>
R:Zain, S.B.; Salim, M.; Chou, W.G.; Sajdel-Sulkowska, E.M.; Majocha, R.E.; Marotta,
Proc. Natl. Acad. Sci. U.S.A. 85, 929-933, 1988
A:Title: Molecular cloning of amyloid cDNA derived from mRNA of the Alzheimer disease
A:Reference number: A31087; MUID:88124954
A:Accession: A31087
A:Molecule type: mRNA
A:Residues: 507-770 <ZAI>
A:Cross-references: GB:M18734; NID:g178572; PIDN:AAA51726.1; PID:g178573
A:Note: the authors translated the codon GAA for residue 599 as Gly, ACC for residue
8 as Val, GTG for residue 609 as Asn, AAT for residue 610 as Gly, and GGT for residue
A:Note: the cited Genbank accession number, J03594, is not in release 101.0
R:Masters, C.L.; Multhaup, G.; Simms, G.; Pottgiesser, J.; Martins, R.N.; Beyreuther,
Query Match 83.7%; Score 41; DB 1; Length 770;
Best Local Similarity 80.0%; Pred. No. 1.8;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 SEVNLDAEFR 10
| | | : | | | | |
Db 667 SEVKMDAEFR 676
| | | : | | | | |
RESULT 10
D42725
nitrite hydratase region 3'-hypothetical protein P47K - Pseudomonas chlororaphis (str
C:Species: Pseudomonas chlororaphis
C>Date: 10-Mar-1994 #sequence_revision 10-Mar-1994 #text_change 20-Jun-2000
C:Accession: D42725
R:Nishiyama, M.; Horinouchi, S.; Kobayashi, M.; Nagasawa, T.; Yamada, H.; Beppu, T.
J. Bacteriol. 173, 2465-2472, 1991
A:Title: Cloning and characterization of genes responsible for metabolism of nitrite
A:Reference number: A42725; MUID:91193202
A:Accession: D42725
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-419 <NIS>
A:Cross-references: GB:D90216; NID:g216850; PIDN:BAAL4247.1; PID:g216854
C:Superfamily: conserved hypothetical protein yciC
Query Match 77.6%; Score 38; DB 2; Length 419;
Best Local Similarity 100.0%; Pred. No. 3.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SEVNLDAE 8
| | | | | | | |
Db 47 SEVNLDAE 54
| | | | | | | |
RESULT 11
A54794
dynein heavy chain, cytosolic - fruit fly (Drosophila melanogaster)
N:Contains: dynein ATPase (EC 3.6.1.33)
C:Species: Drosophila melanogaster
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Jan-2001
C:Accession: A54794
R:Li, M.; McGrail, M.; Serr, M.; Hays, T.S.
J. Cell Biol. 126, 1475-1494, 1994
A:Title: Drosophila cytoplasmic dynein, a microtubule motor that is asymmetrically lo
A:Reference number: A54794; MUID:94375524
A:Accession: A54794
A:Molecule type: mRNA
A:Residues: 1-4639 <LIA>
A:Cross-references: GB:L23195; NID:g349668; PIDN:AAA60323.1; PID:g349669
C:Genetics:
A:Gene: FlyBase:Dnc64C
A:Cross-references: FlyBase:FBgn0010349
C:Superfamily: dynein heavy chain, cytosolic
C:Keywords: Atp; heterotetramer; hydrolase; microtubule binding; nucleotide binding;
F:1895-1902/Region: nucleotide-binding motif A (P-loop)
F:2210-2217/Region: nucleotide-binding motif A (P-loop)

F:2580-2587/Region: nucleotide-binding motif A (P-loop)
 F:2922-2929/Region: nucleotide-binding motif A (P-loop)
 F:1901/Binding site: ATP (Lys) #status predicted
 F:2216/Binding site: ATP (Lys) #status predicted
 F:2596/Binding site: ATP (Lys) #status predicted
 F:2928/Binding site: ATP (Lys) #status predicted

Query Match 77.6% Score 38; DB 1; Length 4639;

Best Local Similarity 88.9% Pred. No. 65;

Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEVNLDAEF 9

|||||

Db 3882 SESNLDAEF 3890

RESULT 12

C75591

threonine synthase - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000

C:Accession: C75591

R.White O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
 S.; Smith, H.O.; Venter, J.C.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896

A:Accession: C75591

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-470 <WHI>

A:Cross-references: GB:AE001863; GB:AE001825; NID:g6460670; PIDN:AAFL2429.1; PID:g646072

A:Experimental source: strain R1

C:Genetics:

A:Gene: DRA0360

A:Map position: 2

C:Superfamily: threonine synthase

Query Match 75.5% Score 37; DB 2; Length 470;

Best Local Similarity 77.8% Pred. No. 7.2;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 EYNLDAEFR 10

|||||

Db 205 EYNADAEFK 213

RESULT 13

S23094

beta-amyloid protein precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 03-May-1996

C:Accession: S23094

R.Kojima, S.; Omori, M.

FEBS Lett. 304, 57-60, 1992

A:Title: Two-way cleavage of beta-amyloid protein precursor by multicatalytic proteinase

A:Reference number: S23094; MUID:92316198

A:Accession: S23094

A:Molecule type: protein.

A:Residues: 1-33 <KOJ>

C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinase

Query Match

Best Local Similarity 73.5% Score 36; DB 2; Length 33;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEVNLDAEF 9

|||||

Db 1 SEVKMDAEF 9

RESULT 14

A27485

Alzheimer's disease amyloid beta/A4 protein homolog precursor - mouse

N:Alternate names: proteinase nexin II

C:Species: Mus musculus (house mouse)

C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 13-Aug-1999

C:Accession: A27485; S19727; I49485

R.Yamada, T.; Sasaki, H.; Furuya, H.; Miyata, T.; Goto, I.; Sakaki, Y.

Biochem. Biophys. Res. Commun. 149, 665-671, 1987

A:Title: Complementary DNA for the mouse homolog of the human amyloid beta protein pr

A:Reference number: A27485; MUID:88106489

A:Accession: A27485

A:Molecule type: mRNA

A:Residues: 1-695 <YAM>

A:Cross-references: GB:M18373; NID:g191568; PIDN:AAA37139.1; PID:g309085

A:Experimental source: brain

R.de Strooper, H.; van Leuven, F.; van den Berghe, H.

Biochim. Biophys. Acta 1129, 141-143, 1991

A:Title: The amyloid beta protein precursor or proteinase nexin II from mouse is clos

A:Reference number: S19727; MUID:92096458

A:Accession: S19727

A:Molecule type: mRNA

A:Residues: 1-210, 'G', 212-220, 'S', 222-396, 'A', 398-402, 'T', 404-448, 'A', 450-695 <STR>

A:Cross-references: EMBL:X59379

R.Izumi, R.; Yamada, T.; Yoshikai, S.; Sasaki, H.; Hattori, M.; Sakaki, Y.

Gene 112, 189-195, 1992

A:Title: Positive and negative regulatory elements for the expression of the Alzheimer

A:Reference number: I49485; MUID:92209598

A:Accession: I49485

A>Status: translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-15 <RES>

A:Cross-references: GB:D10603; NID:g220328; PIDN:BAA01456.1; PID:g220329

C:Genetics:

A:Map position: 16C3

C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinas

C:Keywords: alternative splicing; amyloid; transmembrane protein

Query Match

Best Local Similarity 73.5% Score 36; DB 2; Length 695;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEVNLDAEF 9

|||||

Db 592 SEVKMDAEF 600

RESULT 15

S00550

Alzheimer's disease amyloid beta protein precursor - rat

N:Alternate names: beta-A4 amyloid protein

C:Species: Rattus norvegicus (Norway rat)

C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 13-Aug-1999

C:Accession: S00550; A41245; A39820; S46251

R.Shivers, B.D.; Hilbich, C.; Multhaup, G.; Salbaum, M.; Beyreuther, K.; Seeburg, P.H.

EMBO J. 7, 1365-1370, 1988

A:Title: Alzheimer's disease amyloidogenic glycoprotein: expression pattern in rat br

A:Reference number: S00550; MUID:88312583

A:Accession: S00550

A:Molecule type: mRNA

A:Residues: 1-695 <SHI>

A:Cross-references: EMBL:X07648; NID:g55616; PIDN:CAA30488.1; PID:g55617

R.Schubert, D.; Schroeder, R.; LaCorbiere, M.; Saitoh, T.; Cole, G.

Science 241, 223-226, 1988

A:Title: Amyloid beta protein precursor is possibly a heparan sulfate proteoglycan co

A:Reference number: A41245; MUID:88264430

A:Accession: A41245

A:Molecule type: protein

A:Residues: 18-37, 'X', 39-40, 'X', 42-44 <SCH>

A>Note: evidence for heparan sulfate attachment

R:Hesse, L.; Beher, D.; Masters, C.L.; Multhaup, G.
 FEBS Lett. 349, 109-116, 1994
 A:Title: The beta-A4 amyloid precursor protein binding to copper.
 A:Reference number: S46251; MUID:94320627
 A:Contents: annotation; copper binding sites
 A:Note: rat peptides were isolated but not sequenced
 R:Potempska, A.; Styles, J.; Mehta, P.; Kim, K.S.; Miller, D.L.
 J. Biol. Chem. 266, 8464-8469, 1991
 A:Title: Purification and tissue level of the beta-amyloid peptide precursor of rat brain
 A:Reference number: A39820; MUID:91217087
 A:Accession: A39820
 A>Status: preliminary
 A:Molecule type: protein
 A:Residues: 18-32 <POT>
 A:Experimental source: brain
 C:Comment: Deposition of amyloid protein as neurofibrillary tangles and/or plaques is characteristic of Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinase inhibitor
 C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinase inhibitor
 C:Keywords: alternative splicing; amyloid; glycoprotein; transmembrane protein
 F:625-648/Domain: transmembrane #status predicted <TMM>

Query Match 73.5%; Score 36; DB 2; Length 695;
 Best Local Similarity 77.8%; Pred. No. 19;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEVNLDAEF 9
 III :IIII
 Db 592 SEVKMDAEF 600

Search completed: September 6, 2001, 16:45:48
 Job time: 492 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:51:05 ; Search time 72.75 Seconds
(without alignments)
4.709 Million cell updates/sec

Title: US-09-603-713-5

Perfect score: 49

Sequence: 1 SEVNLDAEFR 10

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	41	83.7	57	1 A4_PIG	Q29023 sus scrofa
2	41	83.7	57	1 A4_URSM	Q29149 ursus marit
3	41	83.7	58	1 A4_CANFA	Q28280 canis famli
4	41	83.7	58	1 A4_RABIT	Q28748 oryctolagus
5	41	83.7	58	1 A4_SHEEP	Q28757 ovis aries
6	41	83.7	59	1 A4_BOVIN	Q28053 bos taurus
7	41	83.7	751	1 A4_SAISC	Q95241 salmtr sci
8	41	83.7	770	1 A4_HUMAN	P05067 homo sapien
9	38	77.6	419	1 P47K_PSECL	P31521 pseudomonas
10	38	77.6	4639	1 DYHC_DROME	P37276 drosophila
11	36	73.5	770	1 A4_MOUSE	P12023 mus musculu
12	36	73.5	770	1 A4_RAT	P08592 rattus norv
13	35	71.4	265	1 YGL7_YEAST	P53133 saccharomyc
14	35	71.4	434	1 YRL4_CAEEL	Q09413 caenorhabdi
15	34	69.4	1024	1 Y075_MYCGE	P47321 mycoplasma
16	33	67.3	279	1 REA2_SCHPO	Q92373 schizosacch
17	33	67.3	345	1 RPOA_BORBU	O51455 borrelia bu
18	32	65.3	197	1 NUOC_RICPR	Q9zdh3 rickettsia
19	32	65.3	433	1 MTH1_YEAST	P35198 saccharomyc
20	32	65.3	493	1 PLTP_MOUSE	P55065 mus musculu
21	32	65.3	848	1 YGLH_HSVEB	P09101 equine herp
22	32	65.3	3075	1 LMA1_HUMAN	P25391 homo sapien
23	31	63.3	68	1 YRB2_HALCU	P17104 halobacteri
24	31	63.3	239	1 HIS4_LACUA	O02131 lactococcus
25	31	63.3	244	1 NOG4_RHIME	P06235 rhizobium m
26	31	63.3	245	1 NODG_RHIME	P06234 rhizobium m
27	31	63.3	245	1 NODG_RHIS3	P72332 rhizobium s
28	31	63.3	246	1 CAH_KLEPN	O52535 klebsiella
29	31	63.3	312	1 IF2B_DROME	P41375 drosophila
30	31	63.3	351	1 YNQ4_YEAST	P53892 saccharomyc
31	31	63.3	368	1 K1CR_XENLA	P08802 xenopus lae
32	31	63.3	693	1 LY54_YEAST	P49367 saccharomyc
33	31	63.3	3562	1 PGCV_CHICK	Q09053 gallus gall

34	30.5	62.2	177	1	IP43_YEAST	P12962 saccharomyc
35	30	61.2	126	1	PF04_CAEEL	Q17435 caenorhabdi
36	30	61.2	241	1	6PGL_TREPA	O83490 treponema p
37	30	61.2	267	1	YBX1_BACSU	P54427 bacillus su
38	30	61.2	303	1	LMB1_CHICK	Q01635 gallus gall
39	30	61.2	326	1	CC14_CAEEL	P18834 caenorhabdi
40	30	61.2	326	1	V772_MYXVL	P29825 myxoma viru
41	30	61.2	346	1	DHAS_MYCSM	P41404 mycobacteri
42	30	61.2	363	1	LACK_AGRRD	Q01937 agrobacteri
43	30	61.2	402	1	RDS1_SCHPO	P53693 schizosacch
44	30	61.2	426	1	RAD9_SCHPO	P26306 schizosacch
45	30	61.2	497	1	GLYA_CHLTR	O84439 chlamydia t

ALIGNMENTS

RESULT 1
A4_PIG
ID A4_PIG STANDARD: PRT: 57 AA.
AC Q29023;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN HOMOLOG [CONTAINS: BETA-AMYLOID
DE PROTEIN (BETA-APP) (A-BETA)] (FRAGMENT).
GN APP.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=92017079; PubMed=1656157;
RA Johnstone E.M., Chaney M.O., Norris F.H., Pascual R., Little S.P.;
RT "Conservation of the sequence of the Alzheimer's disease amyloid
RT peptide in dog, polar bear and five other mammals by cross-species
RT polymerase chain reaction analysis."
RL Brain Res. Mol. Brain Res. 10:299-305(1991).
CC -!- FUNCTION: FUNCTIONAL NEURONAL RECEPTOR WHICH COUPLES TO
CC INTRACELLULAR SIGNALING PATHWAY THROUGH THE GTP-BINDING PROTEIN
CC G(O) (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE APP FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC EMBL; X56127; CAA39592.1;
CC HSSP; P05067; IAML.
DR InterPro; IPR001868;
DR PROSITE; PS00319; A4 EXTRA; PARTIAL.
DR PROSITE; PS00320; A4 INTRA; PARTIAL.
KW Glycoprotein; Amyloid; Neurone; Transmembrane.
FT NON_TER 1
FT CHAIN 6 48 BETA-AMYLOID PROTEIN (POTENTIAL).
FT DOMAIN <1 33 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 34 57 POTENTIAL.
FT NON_TER 57
SQ SEQUENCE 57 AA; 6172 MW; 84209D88EBA82DFA CRC64;

Query Match 83.7%; Score 41; DB 1; Length 57;
Best Local Similarity 80.0%; Pred. No. 0.05;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 SEVNLDAEFR 10

```

Db 1 SEVKMDAEFR 10
|||||
RESULT 2
A4_URSM
ID A4_URSM STANDARD; PRT; 57 AA.
AC Q23149;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN HOMOLOG [CONTAINS: BETA-AMYLOID
DE PROTEIN (BETA-APP) (A-BETA)] (FRAGMENT).
GN APP.
OS Ursus maritimus (Polar bear) (Thalarchos maritimus).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Ursidae; Ursus.
OX NCBI_TaxID=29073;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=92017079; PubMed=1656157;
RA Johnstone E.M., Chaney M.O., Norris F.H., Pascual R., Little S.P.;
RT "Conservation of the sequence of the Alzheimer's disease amyloid
RT peptide in dog, polar bear and five other mammals by cross-species
RT polymerase chain reaction analysis.";
RL Brain Res. Mol. Brain Res. 10:299-305(1991).
CC -1- FUNCTION: FUNCTIONAL NEURONAL RECEPTOR WHICH COUPLES TO
CC INTRACELLULAR SIGNALING PATHWAY THROUGH THE GTP-BINDING PROTEIN
CC G(O) (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE APP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X56125; CAA39590.1; -
CC HSSP; P05067; IAML.
CC InterPro; IPR001868; -
CC PROSITE; PS00319; A4-EXTRA; PARTIAL.
CC PROSITE; PS00320; A4-INTRA; PARTIAL.
CC Glycoprotein; Amyloid; Neurone; Transmembrane.
KW Glycoprotein; Amyloid; Neurone; Transmembrane.
FT NON_TER 1
FT CHAIN 6 48 BETA-AMYLOID PROTEIN (POTENTIAL).
FT DOMAIN <1 33 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 34 57 POTENTIAL.
FT NON_TER 57
FT SEQUENCE 57 AA; 6172 MW; 84209D88EBA82DFA CRC64;
CC -----
DR EMBL; X56128; CAA39593.1; -
DR HSSP; P05067; IAML.
DR InterPro; IPR001868; -
DR PROSITE; PS00319; A4-EXTRA; PARTIAL.
DR PROSITE; PS00320; A4-INTRA; PARTIAL.
KW Glycoprotein; Amyloid; Neurone; Transmembrane.
FT NON_TER 1
FT CHAIN 6 48 BETA-AMYLOID PROTEIN (POTENTIAL).
FT DOMAIN <1 33 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 34 57 POTENTIAL.
FT NON_TER 57
FT SEQUENCE 57 AA; 6172 MW; 84209D88EBA82DFA CRC64;
CC -----
Query Match 83.7%; Score 41; DB 1; Length 57;
Best Local Similarity 80.0%; Pred. No. 0.05;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 SEVNLDAEFR 10
Db 1 SEVKMDAEFR 10
|||||
RESULT 3
A4_CANFA
ID A4_CANFA STANDARD; PRT; 58 AA.
AC Q28280;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN HOMOLOG [CONTAINS: BETA-AMYLOID
DE PROTEIN (BETA-APP) (A-BETA)] (FRAGMENT).
GN APP.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=92017079; PubMed=1656157;
RA Johnstone E.M., Chaney M.O., Norris F.H., Pascual R., Little S.P.;
RT "Conservation of the sequence of the Alzheimer's disease amyloid
RT peptide in dog, polar bear and five other mammals by cross-species
RT polymerase chain reaction analysis.";
RL Brain Res. Mol. Brain Res. 10:299-305(1991).
CC -1- FUNCTION: FUNCTIONAL NEURONAL RECEPTOR WHICH COUPLES TO
CC INTRACELLULAR SIGNALING PATHWAY THROUGH THE GTP-BINDING PROTEIN
CC G(O) (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE APP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X56125; CAA39590.1; -
CC HSSP; P05067; IAML.
CC InterPro; IPR001868; -
CC PROSITE; PS00319; A4-EXTRA; PARTIAL.
CC PROSITE; PS00320; A4-INTRA; PARTIAL.
CC Glycoprotein; Amyloid; Neurone; Transmembrane.
KW Glycoprotein; Amyloid; Neurone; Transmembrane.
FT NON_TER 1
FT CHAIN 6 48 BETA-AMYLOID PROTEIN (POTENTIAL).
FT DOMAIN <1 33 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 34 57 POTENTIAL.
FT NON_TER 57
FT SEQUENCE 58 AA; 6285 MW; 8469D488A2E12DFA CRC64;
CC -----
Query Match 83.7%; Score 41; DB 1; Length 58;
Best Local Similarity 80.0%; Pred. No. 0.051;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 SEVNLDAEFR 10
Db 2 SEVKMDAEFR 11
|||||
RESULT 4
A4_RABIT
ID A4_RABIT STANDARD; PRT; 58 AA.
AC Q28748;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN HOMOLOG [CONTAINS: BETA-AMYLOID
DE PROTEIN (BETA-APP) (A-BETA)] (FRAGMENT).
GN APP.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=92017079; PubMed=1656157;
RA Johnstone E.M., Chaney M.O., Norris F.H., Pascual R., Little S.P.;
RT "Conservation of the sequence of the Alzheimer's disease amyloid
RT peptide in dog, polar bear and five other mammals by cross-species
RT polymerase chain reaction analysis.";
RL Brain Res. Mol. Brain Res. 10:299-305(1991).
CC -1- FUNCTION: FUNCTIONAL NEURONAL RECEPTOR WHICH COUPLES TO
CC INTRACELLULAR SIGNALING PATHWAY THROUGH THE GTP-BINDING PROTEIN
CC G(O) (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE APP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X56125; CAA39590.1; -
CC HSSP; P05067; IAML.
CC InterPro; IPR001868; -
CC PROSITE; PS00319; A4-EXTRA; PARTIAL.
CC PROSITE; PS00320; A4-INTRA; PARTIAL.
CC Glycoprotein; Amyloid; Neurone; Transmembrane.
KW Glycoprotein; Amyloid; Neurone; Transmembrane.
FT NON_TER 1
FT CHAIN 7 49 BETA-AMYLOID PROTEIN (POTENTIAL).
FT DOMAIN <1 34 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 35 58 POTENTIAL.
FT NON_TER 58
FT SEQUENCE 58 AA; 6285 MW; 8469D488A2E12DFA CRC64;
CC -----
Query Match 83.7%; Score 41; DB 1; Length 58;
Best Local Similarity 80.0%; Pred. No. 0.051;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 SEVNLDAEFR 10
Db 2 SEVKMDAEFR 11
|||||

```

```

RL Brain Res. Mol. Brain Res. 10:299-305(1991).
CC -!- FUNCTION: FUNCTIONAL NEURONAL RECEPTOR WHICH COUPLES TO
CC INTRACELLULAR SIGNALING PATHWAY THROUGH THE GTP-BINDING PROTEIN
CC G(O) (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE APP FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X56129; CAA39594.1; -.
DR HSSP; P05067; IAML.
DR InterPro; IPR001868; -.
DR PROSITE; PS00319; A4_EXTRA; PARTIAL.
DR PROSITE; PS00320; A4_INTRA; PARTIAL.
KW Glycoprotein; Amyloid; Neurone; Transmembrane.
FT CHAIN 1 1 BETA-AMYLOID PROTEIN (POTENTIAL).
FT DOMAIN <1 33 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 34 57 POTENTIAL.
FT DOMAIN 58 >58 CYTOPLASMIC (POTENTIAL).
FT NON_TER 58 58
FT SEQUENCE 58 AA; 6300 MW; F434209D88EBA82D CRC64;
DR EMBL; X56129; CAA39594.1; -.
DR HSSP; P05067; IAML.
DR InterPro; IPR001868; -.
DR PROSITE; PS00319; A4_EXTRA; PARTIAL.
DR PROSITE; PS00320; A4_INTRA; PARTIAL.
KW Glycoprotein; Amyloid; Neurone; Transmembrane.
FT CHAIN 1 1 BETA-AMYLOID PROTEIN (POTENTIAL).
FT DOMAIN <1 33 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 34 57 POTENTIAL.
FT DOMAIN 58 >58 CYTOPLASMIC (POTENTIAL).
FT NON_TER 58 58
FT SEQUENCE 58 AA; 6300 MW; F434209D88EBA82D CRC64;

Query Match 83.7%; Score 41; DB 1; Length 58;
Best Local Similarity 80.0%; Pred. No. 0.051;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEVNLDAEFR 10
DB 1 SEVKMDAEFR 10
|||:|||||

RESULT 5
A4_SHEEP
ID A4_SHEEP STANDARD; PRT; 58 AA.
AC Q28757;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN HOMOLOG [CONTAINS: BETA-AMYLOID
DE PROTEIN (BETA-APP) (A-BETA)] (FRAGMENT).
GN APP.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92017079; PubMed=1656157;
RA Johnstone E.M., Chaney M.O., Norris F.H., Pascual R., Little S.P.;
RT "Conservation of the sequence of the Alzheimer's disease amyloid
RT peptide in dog, polar bear and five other mammals by cross-species
RT polymerase chain reaction analysis.";
RL Brain Res. Mol. Brain Res. 10:299-305(1991).
CC -!- FUNCTION: FUNCTIONAL NEURONAL RECEPTOR WHICH COUPLES TO
CC INTRACELLULAR SIGNALING PATHWAY THROUGH THE GTP-BINDING PROTEIN
CC G(O) (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE APP FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X56130; CAA39595.1; -.
DR HSSP; P05057; IAML.
DR InterPro; IPR001868; -.
DR PROSITE; PS00319; A4_EXTRA; PARTIAL.
DR PROSITE; PS00320; A4_INTRA; PARTIAL.
KW Glycoprotein; Amyloid; Neurone; Transmembrane.
FT CHAIN 1 1 BETA-AMYLOID PROTEIN (POTENTIAL).
FT DOMAIN <1 33 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 34 57 POTENTIAL.
FT DOMAIN 58 >58 CYTOPLASMIC (POTENTIAL).
FT NON_TER 58 58
FT SEQUENCE 58 AA; 6300 MW; F434209D88EBA82D CRC64;

Query Match 83.7%; Score 41; DB 1; Length 58;
Best Local Similarity 80.0%; Pred. No. 0.051;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEVNLDAEFR 10
DB 1 SEVKMDAEFR 10
|||:|||||

RESULT 6
A4_BOVIN
ID A4_BOVIN STANDARD; PRT; 59 AA.
AC Q28053;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN HOMOLOG [CONTAINS: BETA-AMYLOID
DE PROTEIN (BETA-APP) (A-BETA)] (FRAGMENT).
GN APP.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92017079; PubMed=1656157;
RA Johnstone E.M., Chaney M.O., Norris F.H., Pascual R., Little S.P.;
RT "Conservation of the sequence of the Alzheimer's disease amyloid
RT peptide in dog, polar bear and five other mammals by cross-species
RT polymerase chain reaction analysis.";
RL Brain Res. Mol. Brain Res. 10:299-305(1991).
CC -!- FUNCTION: FUNCTIONAL NEURONAL RECEPTOR WHICH COUPLES TO
CC INTRACELLULAR SIGNALING PATHWAY THROUGH THE GTP-BINDING PROTEIN
CC G(O) (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE APP FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X56124; CAA39589.1; -.
DR EMBL; X56126; CAA39591.1; -.
DR HSSP; P05067; IAML.
DR InterPro; IPR001868; -.
DR PROSITE; PS00319; A4_EXTRA; PARTIAL.
DR PROSITE; PS00320; A4_INTRA; PARTIAL.
KW Glycoprotein; Amyloid; Neurone; Transmembrane.

```

FT NON_TER 1 1
 FT CHAIN 7 49 BETA-AMYLOID PROTEIN (POTENTIAL).
 FT DOMAIN <1 34 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 35 58 POTENTIAL.
 FT DOMAIN 59 >59 CYTOPLASMIC (POTENTIAL).
 FT NON_TER 59 59
 SQ SEQUENCE 59 AA; 6414 MW; F43469D488A2E12D CRC64;

Query Match 83.7%; Score 41; DB 1; Length 59;
 Best Local Similarity 80.0%; Pred. No. 0.052;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEVNLDAEFR 10
 ||||:|||||
 DB 2 SEVKMDAEFR 11

RESULT 7

AA_SAISC STANDARD; PRT; 751 AA.
 AC Q95241;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN PRECURSOR [CONTAINS: BETA-AMYLOID PROTEIN (BETA-APP) (A-BETA)].
 DE APP.
 GN APP.
 OS Saimiri sciureus (Common squirrel monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Saimiri.
 OC NCBI_TaxID=9521;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver, and kidney;
 RX MEDLINE=96108492; PubMed=85321114;
 RA Levy E., Amorim A., Frangione B., Walker L.C.;
 RT "Beta-amyloid precursor protein gene in squirrel monkeys with cerebral amyloid angiopathy.";
 RL Neurobiol. Aging 16:805-808(1995).
 CC -!- FUNCTION: FUNCTIONAL NEURONAL RECEPTOR WHICH COUPLES TO INTRACELLULAR SIGNALING PATHWAY THROUGH THE GTP-BINDING PROTEIN G(O).
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -!- DOMAIN: THE CLATHRIN-BINDING SITE IS ESSENTIAL FOR ITS ASSOCIATION WITH X11-ALPHA, -BETA, AND -GAMMA. THE SEQUENCE SPECIFIC RECOGNITION EXTENDS TO PEPTIDE RESIDUES THAT ARE C-TERMINAL TO THE NPXY MOTIF. THIS INTERACTION APPEARS TO BE INDEPENDENT OF PHOSPHORYLATION (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE APP FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL; S91024; A014347.1; -
 DR InterPro; IPR001255; -
 DR InterPro; IPR001868; -
 DR InterPro; IPR002223; -
 DR Pfam; PF00014; Kunitz_BPTI; 1.
 DR PRINTS; PR00203; AMYLOIDA4.
 DR PRINTS; PR00204; BETAAMYLOID.
 DR PRINTS; PR00759; BASICPTASE.
 DR PROSITE; PS00319; A4_EXTRA; 1.
 DR PROSITE; PS00320; A4_INTRA; 1.
 DR PROSITE; PS00280; BPTI_KUNITZ.1; 1.
 DR PROSITE; PS00279; BPTI_KUNITZ.2; 1.
 KW Glycoprotein; Amyloid; Neurone; Transmembrane; Alternative splicing; Signal; Serine protease inhibitor.

FT SIGNAL 1 17
 FT CHAIN 18 751 BY SIMILARITY.
 FT CHAIN 653 695 A4 PROTEIN.
 FT DOMAIN 18 680 BETA-AMYLOID PROTEIN (POTENTIAL).
 FT TRANSMEM 681 704 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 705 751 POTENTIAL.
 FT DOMAIN 287 345 CYTOPLASMIC (POTENTIAL).
 FT SITE 740 743 BPTI/KUNITZ INHIBITOR.
 FT SITE 743 743 CLATHRIN-BINDING (BY SIMILARITY).
 FT ACT_SITE 301 302 REACTIVE BOND.
 FT DISULFID 291 341 BY SIMILARITY.
 FT DISULFID 300 324 BY SIMILARITY.
 FT DISULFID 316 337 BY SIMILARITY.
 FT CARBOHYD 523 523 N-LINKED (GLCNAC. . .) (PROBABLE).
 FT CARBOHYD 552 552 N-LINKED (GLCNAC. . .) (PROBABLE).
 SQ SEQUENCE 751 AA; 84893 MW; 6C3E431089569049 CRC64;

Query Match 83.7%; Score 41; DB 1; Length 751;
 Best Local Similarity 80.0%; Pred. No. 0.9;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEVNLDAEFR 10
 ||||:|||||
 DB 648 SEVKMDAEFR 657

RESULT 8

AA_HUMAN STANDARD; PRT; 770 AA.
 ID AC P05067; P09000; Q16011;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN PRECURSOR (PROTEASE NEXIN-II) (PN-II) (APPI) [CONTAINS: BETA-AMYLOID PROTEIN (BETA-APP) (A-BETA)].
 GN APP OR A4 OR CVAP OR ADL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=87144572; PubMed=2881207;
 RA Kang J., Lemaire H.-G., Unterbeck A., Salbaum J.M., Masters C.L., Grzeschik K.-H., Multhaup G., Beyreuther K., Mueller-Hill B.;
 RT "The precursor of Alzheimer's disease amyloid A4 protein resembles a cell-surface receptor.";
 RL Nature 325:733-736(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88122639; PubMed=2893289;
 RA Ponte P., Gonzalez-Dewhitt P., Schilling J., Miller J., Hsu D., Greenberg B., Davis K., Wallace W., Lieberburg I., Fuller F., Cordell B.;
 RT "A new A4 amyloid mRNA contains a domain homologous to serine proteinase inhibitors.";
 RL Nature 331:525-527(1988).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89128427; PubMed=2783775;
 RA Lemaire H.G., Salbaum J.M., Multhaup G., Kang J., Bayney R.M., Unterbeck A., Beyreuther K., Mueller-Hill B.;
 RT "The PreA4(695) precursor protein of Alzheimer's disease A4 amyloid is encoded by 16 exons.";
 RL Nucleic Acids Res. 17:517-522(1989).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97263807; PubMed=9108164;
 RA Hattori M., Tsukahara F., Furuhashi Y., Tanahashi H., Hirose M., Saito M., Tsukuni S., Sakaki Y.;
 RT "A novel method for making nested deletions and its application for sequencing of a 300 kb region of human APP locus.";

RL Nucleic Acids Res. 25:1802-1808(1997).
 RN [5]
 RP SEQUENCE OF 286-345 AND 365-366 FROM N.A.
 RX MEDLINE-88122640; PubMed-2893290;
 RA Tanzi R.E., McClatchey A.I., Lamperti E.D., Villa-Komaroff L.,
 RA Gusella J.F., Neve R.L.;
 RT "Protease inhibitor domain encoded by an amyloid protein precursor
 RT mRNA associated with Alzheimer's disease.";
 RL Nature 331:528-530(1988).
 RN [6]
 RP SEQUENCE OF 287-367 FROM N.A.
 RX MEDLINE-88122641; PubMed-2893291;
 RA Kitaguchi N., Takahashi Y., Tokushima Y., Shiojiri S., Ito H.;
 RT "Novel precursor of Alzheimer's disease amyloid protein shows
 RT protease inhibitory activity.";
 RL Nature 331:530-532(1988).
 RN [7]
 RP SEQUENCE OF 284-289 AND 365-770 FROM N.A.
 RX MEDLINE-87231971; PubMed-3035574;
 RA Robakis N.K., Ramakrishna N., Wolfe G., Wisniewski H.M.;
 RT "Molecular cloning and characterization of a cDNA encoding the
 RT cerebrovascular and the neuritic plaque amyloid peptides.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:4190-4194(1987).
 RN [8]
 RP "Molecular cloning of amyloid cDNA derived from mRNA of the Alzheimer
 RP disease brain: coding and noncoding regions of the fetal precursor
 RA mRNA are expressed in the cortex.";
 RL Proc. Natl. Acad. Sci. U.S.A. 85:929-933(1988).
 RN [9]
 RP SEQUENCE OF 672-681.
 RX MEDLINE-88035004; PubMed-3312495;
 RA Pardridge W.M., Vinters H.V., Yang J., Eisenberg J., Choi T.B.,
 RA Tourtellotte W.W., Huebner V., Shively J.E.;
 RT "Amyloid angiopathy of Alzheimer's disease: amino acid composition
 RT and partial sequence of a 4,200-dalton peptide isolated from cortical
 RT microvessels.";
 RL J. Neurochem. 49:1394-1401(1987).
 RN [10]
 RP SEQUENCE OF 739-770 FROM N.A.
 RX MEDLINE-90236318; PubMed-2110105;
 RA Yoshikai S.-I., Sasaki H., Doh-Ura K., Furuya H., Sakaki Y.;
 RT "Genomic organization of the human amyloid beta-protein precursor
 RT gene.";
 RL Gene 87:257-263(1990).
 RN [11]
 RP SEQUENCE OF 1-10 FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE-89016647; PubMed-3140222;
 RA Schon E.A., Mita S., Sadlock J., Herbert J.;
 RT "A cDNA specifying the human amyloid beta precursor protein (ABPP)
 RT encodes a 95-kDa polypeptide.";
 RL Nucleic Acids Res. 16:9351-9351(1988).
 RN [12]
 RP SEQUENCE OF 18-50.
 RX MEDLINE-87250462; PubMed-3597385;
 RA van Nostrand W.E., Cunningham D.D.;
 RT "Purification of protease nexin II from human fibroblasts.";
 RL J. Biol. Chem. 262:8508-8514(1987).
 RN [13]
 RP IDENTITY OF APP WITH NEXIN-II.
 RX MEDLINE-89384866; PubMed-2506449;
 RA Oltersdorf T., Fritz L.C., Schenk D.B., Lieberburg I.,
 RA Johnson-Wood K.L., Beattie E.C., Ward P.J., Blacher R.W., Dovey H.F.,
 RA Sinha S.;
 RT "The secreted form of the Alzheimer's amyloid precursor protein with
 RT the Kunitz domain is protease nexin-II.";
 RL Nature 341:144-147(1989).
 RN [14]
 RP PROTEASE-SPECIFICITY OF INHIBITOR DOMAIN.

RX MEDLINE-90211252; PubMed-1969731;
 RA Kido H., Fukutomi A., Schilling J., Wang Y., Cordell B., Katunuma N.;
 RT "Protease-specificity of Kunitz inhibitor domain of Alzheimer's
 RL disease amyloid protein precursor.";
 RN Biochem. Biophys. Res. Commun. 167:716-721(1990).
 RN [15]
 RP COMPLEX WITH G(O).
 RX MEDLINE-93189965; PubMed-8446172;
 RA Nishimoto I., Okamoto T., Matsuura Y., Takahashi S., Okamoto T.,
 RA Murayama Y., Ogata E.;
 RT "Alzheimer amyloid protein precursor complexes with brain GTP-binding
 RT protein G(c).";
 RL Nature 362:75-79(1993).
 RN [16]
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 28-133.
 RX MEDLINE-99215582; PubMed-10201399;
 RA Rossjohn J., Cappai R., Feil S.C., Henry A., McKinstry W.J.,
 RA Galatis D., Hesse L., Multhaup G., Beyreuther K., Masters C.L.,
 RA Parker M.W.;
 RT "Crystal structure of the N-terminal, growth factor-like domain of
 RT Alzheimer amyloid precursor protein.";
 RL Nat. Struct. Biol. 6:327-331(1999).
 RN [17]
 RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF 287-344.
 RX MEDLINE-91104913; PubMed-2125487;
 RA Hynes T.R., Randal M., Kennedy L.A., Eigenbrot C., Kossiakof A.A.;
 RT "X-ray crystal structure of the protease inhibitor domain of
 RT Alzheimer's amyloid beta-protein precursor.";
 RL Biochemistry 29:10018-10022(1990).
 RN [18]
 RP STRUCTURE BY NMR OF 289-344.
 RX MEDLINE-92031488; PubMed-1718421;
 RA Heald S.L., Tilton R.F. Jr., Hammond L.S., Lee A., Bayney R.M.,
 RA Kamark M.E., Ramabhadran T.V., Dreyer R.N., Davis G., Unterbeck A.,
 RA Tamburini P.P.;
 RT "Sequential NMR resonance assignment and structure determination of
 RT the Kunitz-type inhibitor domain of the Alzheimer's beta-amyloid
 RT precursor protein.";
 RL Biochemistry 30:10467-10478(1991).
 RN [19]
 RP STRUCTURE BY NMR OF 672-699.
 RX MEDLINE-94281210; PubMed-7516706;
 RA Talafous J., Marcinkowski K.J., Klopman G., Zagorski M.G.;
 RT "Solution structure of residues 1-28 of the amyloid beta-peptide.";
 RL Biochemistry 33:7788-7796(1994).
 RN [20]
 RP STRUCTURE BY NMR OF 696-706.
 RX MEDLINE-97128622; PubMed-8973180;
 RA Kohno T., Kobayashi K., Maeda T., Sato K., Takashima A.;
 RT "Three-dimensional structures of the amyloid beta peptide (25-35) in
 RT membrane-mimicking environment.";
 RL Biochemistry 35:16094-16104(1996).
 RN [21]
 RP STRUCTURE BY NMR OF 672-711.
 RX MEDLINE-98359783; PubMed-9693002;
 RA Coles M., Bicknell W., Watson A.A., Fairlie D.P., Craik D.J.;
 RT "Solution structure of amyloid beta-peptide(1-40) in a water-micelle
 RT environment. Is the membrane-spanning domain where we think it is?";
 RL Biochemistry 37:11064-11077(1998).
 RN [22]
 RP STRUCTURE BY NMR OF 672-699.
 RX MEDLINE-20400066; PubMed-10940222;
 RA Poulsen S.-A., Watson A.A., Craik D.J.;
 RT "Solution structures in aqueous SDS micelles of two amyloid beta
 RT peptides of Abeta(1-28) mutated at the alpha-secretase cleavage
 RT site.";
 RL J. Struct. Biol. 130:142-152(2000).
 RN [23]
 RP SIGNAL SEQUENCE CLEAVAGE SITE, AND TOPOLOGY.
 RX MEDLINE-88296437; PubMed-2900137;
 RA Dyrks T., Weidemann A., Multhaup G., Salbaum J.M., Lemaire H.-G.,
 RA Kang J., Mueller-Hill B., Masters C.L., Beyreuther K.;
 RT "Identification, transmembrane orientation and biogenesis of the

```
RT amyloid A4 precursor of Alzheimer's disease.";
RL EMBO J. 7:949-957(1988).
RN [24]
RP REVIEW.
RX MEDLINE=92271194; PubMed=1589757;
RA Kosik K.S.;
RT "Alzheimer's disease: a cell biological perspective.";
RL Science 256:780-783(1992).
RN [25]

Query Match 83.7%; Score 41; DB 1; Length 770;
Best Local Similarity 80.0%; Pred. No. 0.93;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEVNLDAEFR 10
Db 667 SEVRMDAEFR 676

RESULT . 9
P47K_PSECL
ID P47K_PSECL STANDARD; PRT; 419 AA.
AC P31521;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE 47 KDA PROTEIN (P47K).
OS Pseudomonas chlororaphis (Pseudomonas fluorescens biotype D).
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=333;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B23; PubMed=2013568;
RX MEDLINE=9119202; PubMed=2013568;
RA Nishiyama M., Horinouchi S., Kobayashi M., Nagasawa T., Yamada H.,
RA Beppu T.;
RT "Cloning and characterization of genes responsible for metabolism of
RT nitrile compounds from Pseudomonas chlororaphis B23.";
RL J. Bacteriol. 173:2465-2472(1991).
CC -1- FUNCTION: THE PRESENCE OF P47K IS CRITICAL FOR THE EXPRESSION
CC OF THE NITRILE HYDRATASE GENES. MAY STABILIZE OR ACTIVATE THE
CC NITRILE HYDRATASE PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D90216; BAA14247.1; -
DR PIR; D42725; D42725.
SQ SEQUENCE 419 AA; 46666 MW; FF5113800E27FF0C CRC64;

Query Match 77.6%; Score 38; DB 1; Length 419;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVNLDAAE 8
Db 47 SEVNLDAAE 54

RESULT 10
DYHC_DROME
ID DYHC_DROME STANDARD; PRT; 4639 AA.
AC P37276;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN HOMOLOG
DE (AMYLOIDOGENIC GLYCOPROTEIN) (AG).
GN APP.

DE DYNEIN HEAVY CHAIN, CYTOSOLIC (DYHC).
GN CDHC OR DHC64C.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Phyllozoa; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94375524; PubMed=8089180;
RA Li M., McGrail M., Serr M., Hays T.S.;
RT "Drosophila cytoplasmic dynein, a microtubule motor that is
RT asymmetrically localized in the oocyte.";
RL J. Cell Biol. 126:1475-1494(1994).
RN [2]
RP SEQUENCE OF 1877-1998 FROM N.A.
RX MEDLINE=94243034; PubMed=8186464;
RA Rasmussen K., Serr M., Gepner J., Gibbons I., Hays T.S.;
RT "A family of dynein genes in Drosophila melanogaster.";
RL Mol. Biol. Cell 5:45-55(1994).
CC -1- FUNCTION: DYNEIN HAS ATPASE ACTIVITY. CYTOPLASMIC DYNEIN ACTS AS A
CC MOTOR FOR THE INTRACELLULAR RETROGRADE MOTILITY OF VESICLES AND
CC ORGANELLES ALONG MICROTUBULES.
CC -1- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF
CC INTERMEDIATE AND LIGHT CHAINS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L23195; AAA60323.1; -
DR EMBL; L25122; AAA28492.1; -
DR FlyBase; FBgn0010349; Dhc64C.
KW Motor protein; Microtubules;
FT DOMAIN 530 565
FT DOMAIN 774 794
FT DOMAIN 1264 1368
FT DOMAIN 1999 2027
FT DOMAIN 3189 3261
FT DOMAIN 3382 3478
FT DOMAIN 3723 3782
FT DOMAIN 1895 1902
FT NP_BIND 2210 2217
FT NP_BIND 2580 2587
FT NP_BIND 2922 2929
SQ SEQUENCE 4639 AA; 530152 MW; 057A7D800CCD07E CRC64;

Query Match 77.6%; Score 38; DB 1; Length 4639;
Best Local Similarity 88.9%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEVNLDAEFR 9
Db 3882 SESNLDAEFR 3890

RESULT 11
A4_MOUSE
ID A4_MOUSE STANDARD; PRT; 770 AA.
AC P12023;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN HOMOLOG
DE (AMYLOIDOGENIC GLYCOPROTEIN) (AG).
GN APP.
```


OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE OF 1-289 AND 365-770 FROM N.A.
RC STRAIN=BAUB/C; TISSUE=Brain;
RX MEDLINE=92096458; PubMed=1756177;
RA de Strooper B., van Leuven F., van den Berghe H.;
RT "The amyloid beta protein precursor or proteinase nexin II from mouse
RL is closer related to its human homolog than previously reported.";
RN Biochim. Biophys. Acta 1129:141-143(1991).
RN [2]
RP SEQUENCE OF 1-289 AND 365-770 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=88106489; PubMed=3322280;
RA Yamada T., Sakaki H., Furuya H., Miyata T., Goto I., Sakaki Y.;
RT "Complementary DNA for the mouse homolog of the human amyloid beta
RL protein precursor.";
RN Biochem. Biophys. Res. Commun. 149:665-671(1987).
RN [3]
RP REVISIONS.
RA Yamada T.;
RL Submitted (MAR-1988) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 289-364 FROM N.A.
RC STRAIN=CD-1; TISSUE=Placenta;
RX MEDLINE=89345111; PubMed=25697110;
RA Fukuchi K., Martin G.M., Deeb S.S.;
RT "Sequence of the protease inhibitor domain of the A4 amyloid protein
RL precursor of Mus domesticus.";
RN Nucleic Acids Res. 17:5396-5396(1989).
RN [5]
RP SEQUENCE OF 1-19 FROM N.A.
RX MEDLINE=92209998; PubMed=1555768;
RA Izumi R., Yamada T., Yoshikai S.I., Sasaki H., Hattori M.,
RA Sakai Y.;
RT "Positive and negative regulatory elements for the expression of the
RL Alzheimer's disease amyloid precursor-encoding gene in mouse.";
RN Gene 112:189-195(1992).
RN [6]
RP SEQUENCE OF 281-380 FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=Brain, and Kidney;
RX MEDLINE=89149813; PubMed=2493250;
RA Yamada T., Sakaki H., Dohura K., Goto I., Sakaki Y.;
RT "Structure and expression of the alternatively-spliced forms of mRNA
RL for the mouse homolog of Alzheimer's disease amyloid beta protein
RT precursor.";
RL Biochem. Biophys. Res. Commun. 158:906-912(1989).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: SIX FORMS OF APP ARE FOUND; APP(395),
CC APP(563), APP(695), APP(751) AND APP(770) (SHOWN HERE); WHICH ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: AAA(770) IS EXPRESSED IN KIDNEY. AAA(751) IS
CC WIDELY EXPRESSED. AAA(695) IS EXPRESSED IN BRAIN, KIDNEY AND
CC LIVER.
CC -1- DOMAIN: THE CLATHRIN-BINDING SITE IS ESSENTIAL FOR ITS ASSOCIATION
CC WITH X11-ALPHA, -BETA, AND -GAMMA. THE SEQUENCE SPECIFIC
CC RECOGNITION EXTENDS TO PEPTIDE RESIDUES THAT ARE C-TERMINAL TO THE
CC NPXY MOTIF. THIS INTERACTION APPEARS TO BE INDEPENDENT OF
CC PHOSPHORYLATION (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE APP FAMILY.
CC -1- SIMILARITY: APP(751) AND APP(770) CONTAIN A PROTEASE INHIBITOR
CC DOMAIN BELONGING TO THE BPTI/KUNITZ FAMILY OF INHIBITORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; X59379; -; NOT_ANNOTATED_CDS.
DR EMBL; M18373; AAA37139.1; -.
DR EMBL; X15210; CAA33280.1; -.
DR EMBL; D10603; BAA01456.1; -.
DR EMBL; M24397; AAA39929.1; -.
DR PIR; A27485; A27485.
DR PIR; S04855; S04855.
DR PIR; S19727; S19727.
DR MGD; MGI:88059; App.
DR InterPro; IPR001255; -.
DR InterPro; IPR001868; -.
DR InterPro; IPR002223; -.
DR Pfam; PF00014; Kunitz_BPTI; 1.
DR PRINTS; PR00203; AMYLOIDA4.
DR PRINTS; PR00204; BETAAMYLOID.
DR PRINTS; PR00759; BASICPTASE.
DR PROSITE; PS00319; A4_EXTRA; 1.
DR PROSITE; PS00320; A4_INTRA; 1.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 1.
DR PROSITE; PS00279; BPTI_KUNITZ_2; 1.
KW Glycoprotein; Amyloid; Neurone; Transmembrane; Signal;
KW Alternative splicing; Serine protease inhibitor.
FT SIGNAL 1 17
FT CHAIN 18 770
FT ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN
FT HOMOLOG.
FT DOMAIN 18 699
FT TRANSMEM 700 723
FT DOMAIN 724 770
FT DOMAIN 673 715
FT DOMAIN 287 345
FT SITE 759 762
FT DISULFID 291 341
FT DISULFID 300 324
FT DISULFID 316 337
FT CARBOHYD 542 542
FT CARBOHYD 571 571
FT VARSPPLIC 269 289
FT VARSPPLIC 290 364
FT VARSPPLIC 346 380
SQ SEQUENCE 770 AA; 86752 MW; 26C50DE0890CAF7A CRC64;

Query Match 73.5%; Score 36; DB 1; Length 770;
Best Local Similarity 77.8%; Pred. No. 10;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEVNLDAEF 9
Db 667 SEVKMDAEF 675
|||:||||

RESULT 12
A4_RAT
ID A4_RAT STANDARD; PRT; 770 AA.
AC P08592;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN HOMOLOG PRECURSOR
DE (AMYLOIDOGENIC GLYCOPROTEIN) (AG).
GN APP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE OF 1-289 AND 365-770 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=88312583; PubMed=2900758;
RA Shivers B.D., Hilbich C., Multhaup G., Salbaum J.M., Beyreuther K.,
RA Seeburg P.E.;
RT "Alzheimer's disease amyloidogenic glycoprotein: expression pattern
RT in rat brain suggests a role in cell contact.";

EMBO J. 7:1365-1370(1988).

[2]
SEQUENCE OF 289-364 FROM N.A.
TISSUE=Liver;
Kang J., Mueller-Hill B.;
"The sequence of the two extra exons in rat preA4.";
Nucleic Acids Res. 17:2130-2130(1989).
-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
-1- ALTERNATIVE PRODUCTS: SIX FORMS OF APP ARE FOUND; APP(395),
APP(563), APP(695), APP(751) AND APP(770) (SHOWN HERE); WHICH ARE
PRODUCED BY ALTERNATIVE SPLICING.
-1- DOMAIN: THE CLATHRIN-BINDING SITE IS ESSENTIAL FOR ITS ASSOCIATION
WITH X11-ALPHA, -BETA, AND -GAMMA. THE SEQUENCE SPECIFIC
RECOGNITION EXTENDS TO PEPTIDE RESIDUES THAT ARE C-TERMINAL TO THE
NPXY MOTIF. THIS INTERACTION APPEARS TO BE INDEPENDENT OF
PHOSPHORYLATION (BY SIMILARITY).
-1- SIMILARITY: BELONGS TO THE APP FAMILY.
-1- SIMILARITY: APP(751) AND APP(770) CONTAIN A PROTEASE INHIBITOR
DOMAIN BELONGING TO THE BPTI/KUNITZ FAMILY OF INHIBITORS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; X07648; CAA30488.1; -;
DR EMBL; X14066; CAA32229.1; -;
DR PIR; S00550; S00550.
DR InterPro; IPR001255; -;
DR InterPro; IPR001868; -;
DR InterPro; IPR002223; -;
DR Pfam; PF00014; Kunitz_BPTI; 1.
DR PRINTS; PR00203; AMYLOIDA4.
DR PRINTS; PR00204; BETAAMYLOID.
DR PRINTS; PR00759; BASICPTASE.
DR PROSITE; PS00319; A4-EXTRA; 1.
DR PROSITE; PS00320; A4-INTRA; 1.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 1.
DR PROSITE; PS02079; BPTI_KUNITZ_2; 1.
DR Glycoprotein; Amyloid; Neurone; Transmembrane; signal;
KW Alternative splicing; Serine protease inhibitor.
FT SIGNAL 1 17
FT CHAIN 18 770
FT
FT DOMAIN 18 699
FT TRANSMEM 700 723
FT DOMAIN 724 770
FT DOMAIN 773 715
FT DOMAIN 287 345
FT SITE 759 762
FT DISULFID 291 341
FT DISULFID 300 324
FT DISULFID 316 337
FT CARBOHYD 542 542
FT CARBOHYD 571 571
FT VARSPPLIC 289 289
FT VARSPPLIC 290 364
SQ SEQUENCE 770 AA; 86704 MW; C26C9D6B2D929A7 CRC64;

Query Match 73.5%; Score 36; DB 1; Length 770;
Best Local Similarity 77.8%; Pred. No. 10;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SEVNLDAEF 9
|||:||||
Db 667 SEVKMDAEF 675

RESULT 13
YGL7_YEAST
ID YGL7_YEAST STANDARD; PRT; 265 AA.
AC P53133;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOTHETICAL 30.8 KDA PROTEIN IN ABC1-CDC20 INTERGENIC REGION.
GN YGL117W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Lauquin G.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; Z72639; CAA96825.1; -;
DR SGD; S0003085; YGL117W.
KW Hypothetical protein.
SQ SEQUENCE 265 AA; 30755 MW; 7BB5E25246CEB75F CRC64;

Query Match 71.4%; Score 35; DB 1; Length 265;
Best Local Similarity 77.8%; Pred. No. 4.9;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SEVNLDAEF 9
|||:||||
Db 238 SDVNLDFEF 246

RESULT 14
YRL4_CAEEL
ID YRL4_CAEEL STANDARD; PRT; 434 AA.
AC Q09413;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOTHETICAL 49.2 KDA PROTEIN R03D7.4 IN CHROMOSOME II.
GN R03D7.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Swinburne J.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; Z46828; CAA86857.1; -;
DR WormPep; R03D7.4; CE01611.
KW Hypothetical protein.
FT DOMAIN 154 159
POLY-SER.

FT DOMAIN 407 414 POLY-ASN.
SQ SEQUENCE 434 AA; 49240 MW; EB0A02A996ED0844 CRC64;

Query Match 71.4%; Score 35; DB 1; Length 434;
Best Local Similarity 87.5%; Pred. No. 8.6;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVNLDADF 9
||| ||| ||
Db 97 EVNLDDEF 104

RESULT 15

Y075_MYCGE
ID Y075_MYCGE STANDARD; PRT: 1024 AA.
AC P47321: O49190;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL PROTEIN MG075.
GN MG075.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
RN [2]
RP SEQUENCE OF 1-156; 269-402; 643-736 AND 808-947 FROM N.A.
RC STRAIN-ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
sequencing.";
RL J. Bacteriol. 175:7918-7930(1993).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U39688; AAC71293.1;
DR EMBL; U01715; AAC43189.1; ALT_INIT.
DR EMBL; U02251; AAD12514.1;
DR EMBL; U01749; AAD10562.1;
DR EMBL; U01775; AAD10595.1;
DR TIGR; MG075;
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 11 31 POTENTIAL.
FT TRANSMEM 951 971 POTENTIAL.
SQ SEQUENCE 1024 AA; 116424 MW; 5B1540E6855CB554 CRC64;

Query Match 69.4%; Score 34; DB 1; Length 1024;
Best Local Similarity 77.8%; Pred. No. 36;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEVNLDADF 9
||| ||| ||

Db 150 SEVSLDLEF 158

Search completed: September 6, 2001, 16:51:05
Job time: 809 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:49:46 ; Search time 231.42 Seconds
(without alignments)
5.717 Million cell updates/sec

Title: US-09-603-713-5
Perfect score: 49
Sequence: 1 SEVNLDAEFR 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 13230527 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_16:.*
1: sp_archaea:.*
2: sp_bacteria:.*
3: sp_fungi:.*
4: sp_human:.*
5: sp_invertebrate:.*
6: sp_mammal:.*
7: sp_mhc:.*
8: sp_organelle:.*
9: sp_phage:.*
10: sp_plant:.*
11: sp_rodent:.*
12: sp_unclassified:.*
13: sp_vertebrate:.*
14: sp_virus:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	83.7	82	4	P78438 homo sapien
2	41	83.7	82	4	Q16014 homo sapien
3	41	83.7	82	4	Q16019 homo sapien
4	41	83.7	82	4	Q16020 homo sapien
5	41	83.7	534	13	Q93296 gallus gall
6	41	83.7	569	13	Q9PVU1 gallus gall
7	41	83.7	695	13	Q60496 cavia porce
8	41	83.7	695	13	Q9DGJ8 gallus gall
9	41	83.7	751	13	Q9DGJ7 gallus gall
10	41	83.7	770	6	Q9PUI0 sus scrofa
11	38	77.6	779	5	Q9WIB1 drosophila
12	38	77.6	4623	5	Q9VZ83 drosophila
13	37	75.5	470	2	Q9RYF8 drosophila
14	36	73.5	79	11	Q35463 cricetus
15	36	73.5	695	11	P97487 mus musculus
16	35	71.4	282	5	O02335 caenorhabdi
17	35	71.4	747	13	Q91963 xenopus lae
18	35	71.4	929	3	Q9HF19 neurospora
19	34	69.4	302	9	Q37840 bacterioph

20	34	69.4	423	2	Q45693 burkholderi
21	34	69.4	423	2	O52379 pseudomonas
22	34	69.4	549	5	Q9V6S5 drosophila
23	34	69.4	4472	5	Q9VDC0 drosophila
24	33	67.3	52	10	Q9M3F9 arabidopsis
25	33	67.3	143	4	Q9H935 homo sapien
26	33	67.3	269	2	O52512 streptomyce
27	33	67.3	442	2	Q9PCF1 xylella fas
28	33	67.3	539	4	Q9NW45 homo sapien
29	33	67.3	539	4	Q9NWD6 homo sapien
30	33	67.3	1181	5	Q9V795 drosophila
31	33	67.3	2514	5	Q9Y061 caenorhabdi
32	33	67.3	2531	5	Q92258 caenorhabdi
33	33	65.3	103	2	O07775 mycobacteri
34	32	65.3	132	3	P78948 schizosacch
35	32	65.3	132	3	Q9P6S6 schizosacch
36	32	65.3	142	5	O16896 caenorhabdi
37	32	65.3	239	10	Q9FNG2 arabidopsis
38	32	65.3	245	2	Q9KJVB rhizobium l
39	32	65.3	313	2	Q9RUL3 deinococcus
40	32	65.3	347	13	Q9YI63 gallus gall
41	32	65.3	347	13	Q9PUU8 gallus gall
42	32	65.3	383	2	Q9S5V6 bacillus ps
43	32	65.3	672	3	Q04562 saccharomyc
44	32	65.3	672	10	Q9ZWB1 arabidopsis
45	32	65.3	692	10	Q9SAH3 arabidopsis

ALIGNMENTS

RESULT	1
P78438	PRELIMINARY;
ID	P78438
AC	P78438;
DT	01-MAY-1997 (TREMBLrel. 03, Created)
DT	01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE	AMYLOID PROTEIN (BETA-AMYLOID PROTEIN) (FRAGMENT).
GN	APP.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=89392030; PubMed=2675837;
RA	Johnstone R.M., Chaney M.O., Moore R.E., Ward K.E., Norris F.H., Little S.P.;
RT	"Alzheimer's disease amyloid peptide is encoded by two exons and shows similarity to soybean trypsin inhibitor.";
RL	Biochem. Biophys. Res. Commun. 163:1248-1255(1989).
RN	[2]
RP	SEQUENCE OF 19-48 FROM N.A.
RX	MEDLINE=87120329; PubMed=2949367;
RA	Tanzi R.E., Gusella J.F., Watkins P.C., Bruns G.A., George-Hyslop P., Van Keuren M.L., Patterson D., Pagan S., Kurnit D.M., Neve R.L.;
RT	"Amyloid beta protein gene: cDNA, mRNA distribution, and genetic linkage near the Alzheimer locus.";
RL	Science 235:880-884(1987).
RN	[3]
RP	SEQUENCE OF 32-63 FROM N.A.
RX	MEDLINE=93035397; PubMed=1415269;
RA	Kamino K., Orr H.T., Payami H., Wijsman E.M., Alonso M.E., Pulst S.M., Anderson L., O'dahl S., Nemens E., White J.A.;
RT	"Linkage and mutational analysis of familial Alzheimer disease kindreds for the APP gene region.";
RL	Am. J. Hum. Genet. 51:998-1014(1992).
DR	EMBL; M29270; AAA51768.1;
DR	EMBL; M29269; AAA51768.1; JOINED.
DR	EMBL; M15532; AAA51564.1;
DR	EMBL; S45136; AAB23646.1;
DR	HSSP; P05067; 1BA4.

Fri Sep 7 10:58:40 2001

us-09-603-713-5.rspt

FT NON_TER 1 1
SQ SEQUENCE 82 AA; 8994 MW; 8DA9EA42B813A070E CRC64;

Query Match 83.7%; Score 41; DB 4; Length 82;
Best Local Similarity 80.0%; Pred. No. 0.55;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEVNLDAEFR 10
|||:|||||
Db 12 SEVKMDAEFR 21

RESULT 2

ID Q16014 PRELIMINARY; PRT; 82 AA.
AC Q16014;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE BETA-AMYLLOID PEPTIDE PRECURSOR (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93236601; PubMed=8476439;
RA Denman R.B., Rosenczwaig R., Miller D.L.;
RT "A system for studying the effect(s) of familial Alzheimer disease mutations on the processing of the beta-amyloid peptide precursor.";
RL Biochem. Biophys. Res. Commun. 192:96-103(1993).
DR EMBL; S60721; AAB26263.2; -
DR HSSP; P05067; 1BA4.
FT NON_TER 1 1
FT NON_TER 82 82
SQ SEQUENCE 82 AA; 8972 MW; F534AA5B3EA9230A CRC64;

Query Match 83.7%; Score 41; DB 4; Length 82;
Best Local Similarity 80.0%; Pred. No. 0.55;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEVNLDAEFR 10
|||:|||||
Db 13 SEVKMDAEFR 22

RESULT 3

ID Q16019 PRELIMINARY; PRT; 82 AA.
AC Q16019;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE BETA-AMYLLOID PEPTIDE PRECURSOR (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93236601; PubMed=8476439;
RA Denman R.B., Rosenczwaig R., Miller D.L.;
RT "A system for studying the effect(s) of familial Alzheimer disease mutations on the processing of the beta-amyloid peptide precursor.";
RL Biochem. Biophys. Res. Commun. 192:96-103(1993).
DR EMBL; S61380; AAB26264.2; -
DR HSSP; P05067; 1BA4.
FT NON_TER 1 1
FT NON_TER 82 82
SQ SEQUENCE 82 AA; 8938 MW; F534AA50E579230A CRC64;

Query Match 83.7%; Score 41; DB 4; Length 82;
Best Local Similarity 80.0%; Pred. No. 0.55;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEVNLDAEFR 10
|||:|||||
Db 13 SEVKMDAEFR 22

RESULT 4

ID Q16020 PRELIMINARY; PRT; 82 AA.
AC Q16020;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE BETA-AMYLLOID PEPTIDE PRECURSOR (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93236601; PubMed=8476439;
RA Denman R.B., Rosenczwaig R., Miller D.L.;
RT "A system for studying the effect(s) of familial Alzheimer disease mutations on the processing of the beta-amyloid peptide precursor.";
RL Biochem. Biophys. Res. Commun. 192:96-103(1993).
DR EMBL; S61383; AAB26265.2; -
DR HSSP; P05067; 1BA4.
FT NON_TER 1 1
FT NON_TER 82 82
SQ SEQUENCE 82 AA; 8882 MW; F534AA5AE5D9230A CRC64;

Query Match 83.7%; Score 41; DB 4; Length 82;
Best Local Similarity 80.0%; Pred. No. 0.55;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEVNLDAEFR 10
|||:|||||
Db 13 SEVKMDAEFR 22

RESULT 5

ID Q93296 PRELIMINARY; PRT; 534 AA.
AC Q93296;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE AMYLLOID PRECURSOR PROTEIN (FRAGMENT).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Barnes N.Y., Ling L., Yoshikawa K., Schwartz L.M., Oppenheim R.W.,
RT "Increased production of amyloid precursor protein provides a substrate for Caspase 3 in dying motoneurons.";
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF042098; AAC25052.1; -
DR HSSP; P05067; 1BA4.
FT NON_TER 1 1
FT NON_TER 82 82
SQ SEQUENCE 82 AA; 8882 MW; F534AA50E579230A CRC64;

```
FT NON_TER 1 1
SQ SEQUENCE 534 AA; 60597 MW; FB53ECC2E6D4C92 CRC64;

Query Match 83.7%; Score 41; DB 13; Length 534;
Best Local Similarity 80.0%; Pred. No. 4.2;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SEVNLDAEFR 10
   ||| :||||
Db 431 SEVKMDAEFR 440

RESULT 6
Q9PVL1
ID Q9PVL1 PRELIMINARY; PRT; 569 AA.
AC Q9PVL1;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE AMYLOID PROTEIN PRECURSOR (FRAGMENT).
GN APP.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Coulson E.J., Paliga K., Beyreuther K., Masters C.L.;
RT "What the evolution of the amyloid protein precursor supergene family
RT tells us about its function."
RL Neurochem. Int. 0:0-0(2000).
DR EMBL; AF030341; AAF12698.1; -.
DR HSSP; P05067; 1BA4.
DR InterPro; IPR001868; -.
DR Pfam; PF02177; A4_EXTRA; 1.
DR PRINTS; PR00203; AMYLOIDA4.
DR PROSITE; PS00319; A4_EXTRA; 1.
DR PROSITE; PS00320; A4_INTRA; 1.
DR SMART; SM00006; A4_EXTRA; 1.
FT NON_TER 1 1
SQ SEQUENCE 569 AA; 64753 MW; 0AB8BB851863A19D CRC64;

Query Match 83.7%; Score 41; DB 13; Length 569;
Best Local Similarity 80.0%; Pred. No. 4.5;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SEVNLDAEFR 10
   ||| :||||
Db 467 SEVKMDAEFR 476

RESULT 7
Q60496
ID Q60496 PRELIMINARY; PRT; 695 AA.
AC Q60496;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Beck M., Mueller D., Bigl V.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: FUNCTIONAL NEURONAL RECEPTOR WHICH COUPLES TO
```

```
CC CC INTRACELLULAR SIGNALING PATHWAY THROUGH THE GTP-BINDING PROTEIN
CC G(O).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- SIMILARITY: TO OTHER SPECIES APP ANALOGUES.
DR EMBL; X97631; CAA66230.1; -.
DR HSSP; P05067; 1BA4.
DR InterPro; IPR001868; -.
DR Pfam; PF02177; A4_EXTRA; 1.
DR PRINTS; PR00203; AMYLOIDA4.
DR PROSITE; PS00319; A4_EXTRA; 1.
DR PROSITE; PS00320; A4_INTRA; 1.
DR SMART; SM00006; A4_EXTRA; 1.
SQ SEQUENCE 695 AA; 78701 MW; 5196A0C4017F16AB CRC64;

Query Match 83.7%; Score 41; DB 11; Length 695;
Best Local Similarity 80.0%; Pred. No. 5.6;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SEVNLDAEFR 10
   ||| :||||
Db 592 SEVKMDAEFR 601

RESULT 8
Q9DGJ8
ID Q9DGJ8 PRELIMINARY; PRT; 695 AA.
AC Q9DGJ8;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE BETA-AMYLOID PRECURSOR PROTEIN 695 ISOFORM.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Sarasa M., Rodolosse A., Sorribas V.;
RT "Cloning of full-length chicken beta-amyloid precursor protein
RT isoforms."
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF289218; AAG00593.1; -.
SQ SEQUENCE 695 AA; 78565 MW; F201ED02AEC86D95 CRC64;

Query Match 83.7%; Score 41; DB 13; Length 695;
Best Local Similarity 80.0%; Pred. No. 5.6;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SEVNLDAEFR 10
   ||| :||||
Db 592 SEVKMDAEFR 601

RESULT 9
Q9DGJ7
ID Q9DGJ7 PRELIMINARY; PRT; 751 AA.
AC Q9DGJ7;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE BETA-AMYLOID PRECURSOR PROTEIN 751 ISOFORM.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Sarasa M., Rodolosse A., Sorribas V.;
RT "Cloning of full-length chicken beta-amyloid precursor protein
```

```
RT isoforms."
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF289219; AAC00594.1; E78E9413A8033D84 CRC64;
SQ SEQUENCE 751 AA; 84705 MW; 87705 MW; 87705 MW; 87705 MW;

Query Match 83.7%; Score 41; DB 13; Length 751;
Best Local Similarity 80.0%; Pred. No. 6.1;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SEVNLDAEFR 10
   ||| :||||
Db 648 SEVKMDAEFR 676

RESULT 10
Q9TU10 PRELIMINARY; PRT; 770 AA.
AC Q9TU10;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE ANYLOID PRECURSOR PROTEIN.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA Kimura A., Takahashi T.;
RT "Anyloid precursor protein 770."
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
DR EMBL: AB032550; BAA84580.1;
DR HSSP; P05067; 1AAP.
DR InterPro; IPR001868;
DR InterPro; IPR002223;
DR Pfam; PF00014; Kunitz_BPTI; 1.
DR Pfam; PF02177; A4_EXTRA; 1.
DR PRINTS; PR00203; ANYLOIDA4.
DR PRINTS; PR00759; BASICPTASE.
DR PROSITE; PS00319; A4_EXTRA; 1.
DR PROSITE; PS00320; A4_INTRA; 1.
DR PROSITE; PS00280; BPTI_KUNITZ; 1.
DR SMART; SM00006; A4_EXTRA; 1.
KW Serine protease inhibitor.
SQ SEQUENCE 770 AA; 86361 MW; 5F7A1DCB2BCC583E CRC64;

Query Match 83.7%; Score 41; DB 6; Length 770;
Best Local Similarity 80.0%; Pred. No. 6.3;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SEVNLDAEFR 10
   ||| :||||
Db 667 SEVKMDAEFR 676

RESULT 11
Q9W1B1 PRELIMINARY; PRT; 779 AA.
AC Q9W1B1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE CG16786 PROTEIN.
DE CG16786.
GN Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=BERKELEY;
MEDLINE=20196006; PubMed=10731132;
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL: AE003463; AA547161.1;
DR FlyBase; FBgn0034974; CG16786.
SQ SEQUENCE 779 AA; 86335 MW; E1231D45CB01A15E CRC64;

Query Match 77.6%; Score 38; DB 5; Length 779;
Best Local Similarity 77.8%; Pred. No. 26;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SEVNLDAEFR 9
   ||| :|||
Db 570 SEINLDLEF 578

RESULT 12
Q9VZ83 PRELIMINARY; PRT; 4623 AA.
AC Q9VZ83;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE DHC64C PROTEIN.
GN DHC64C OR CG7507.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=BERKELEY;
MEDLINE=20196006; PubMed=10731132;
```


RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananides P.G., Scherf S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burtis K.C., Buser D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cavley S., Dahle C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson R.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 DR EMBL; AE003482; AAF47942.2; -
 DR FlyBase; FBgn010349; Dhc64C.
 DR InterPro; IPR000169; -
 DR InterPro; IPR001064; -
 DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
 DR PROSITE; PS00639; THIOLEPROTEASE_HIS; UNKNOWN_1.
 SQ SEQUENCE 4623 AA; 528403 MW; 0E03827AA75818D0 CRC64;

Query Match 77.6%; Score 38; DB 5; Length 4623;
 Best Local Similarity 88.9%; Pred. No. 1.8e+02;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 SEVNLDAAEF 9
 Db 3866 SESNLDAAEF 3874

RESULT 13
 Q9RYF8 PRELIMINARY; PRT; 470 AA.
 ID Q9RYF8
 AC Q9RYF8:
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE THREONINE SYNTHASE.
 CN DRA0360.
 OS Deinococcus radiodurans.
 OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
 OX NCBI_TaxID=1299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=R1.
 RX MEDLINE=20036896; PubMed=10567266;
 RA White O., Eissen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,

RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 RA Wofat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
 RA Vanthaeven J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 RA Fraser C.M.;
 RT "Genome sequence of the radioresistant bacterium Deinococcus
 RT radiodurans R1";
 RL Science 286:1571-1577(1999).
 DR EMBL; AE001863; AAF12429.1; -
 DR TIGR; DRA0360; -
 DR InterPro; IPR000634; -
 DR InterPro; IPR001926; -
 DR Pfam; PF00291; PALP; 1.
 DR PROSITE; PS00165; DEHYDRATASE_SER_THR; 1.
 SQ SEQUENCE 470 AA; 51131 MW; 124A0F0C6341421A CRC64;

Query Match 75.5%; Score 37; DB 2; Length 470;
 Best Local Similarity 77.8%; Pred. No. 24;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 2 EVNLDAAEF 10
 Db 205 EVNLDAAEF 213

RESULT 14
 Q35463 PRELIMINARY; PRT; 79 AA.
 ID Q35463
 AC Q35463:
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE ALZHEIMER'S AMYLOID BETA PROTEIN PRECURSOR (FRAGMENT).
 GN BETA APP.
 OS Cricetus griseus (Chinese hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Cricetulus.
 OX NCBI_TaxID=10029;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Sambamurti K., Pinnix I., Gandhi S.;
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF030413; AAB86608.1; -
 DR HSP; P05067; IQCM.
 FT NON_TER 1
 FT NON_TER 79
 SQ SEQUENCE 79 AA; 8538 MW; 37F2C6C3BFF3F597 CRC64;

Query Match 73.5%; Score 36; DB 11; Length 79;
 Best Local Similarity 77.8%; Pred. No. 5.4;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 SEVNLDAAEF 9
 Db 16 SEVNLDAAEF 24

RESULT 15
 P97487 PRELIMINARY; PRT; 695 AA.
 ID P97487
 AC P97487: P97942;
 DT 01-MAY-1997 (TREMBLrel. 03, Created)
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE HIPPOCAMPAL AMYLOID PROTEIN.
 GN APP.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SAMP8; TISSUE=HIPPOCAMPUS;
 RA Flood J.F., Kumar V.B., Sasser T., Word I., Morley J.E.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 581-662 FROM N.A.
 RC STRAIN=129SV;
 RA Wragg M.A., Busfield F., Duff K., Korenblat K., Capecchi M.,
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U84012; AAB41502.1; -.
 DR EMBL; U82624; AAB40919.1; -.
 DR HSSP; P05067; IQCM
 DR InterPro; IPR001868; -.
 DR Pfam; PF02177; A4_EXTRA; 1.
 DR PRINTS; PR00203; AMYLOIDA4.
 DR PROSITE; PS00319; A4_EXTRA; 1.
 DR PROSITE; PS00320; A4_INTRA; 1.
 DR SMART; SM00006; A4_EXTRA; 1.
 SQ SEQUENCE 695 AA; 78414 MW; 9A5FBE2ED261236E CRC64;

Query Match 73.5%; Score 36; DB 11; Length 695;
 Best Local Similarity 77.8%; Pred. No. 58;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SEVNLDAEF 9
 Db 592 SEVRMDAEF 600

Search completed: September 6, 2001, 16:49:46
 Job time: 730 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:49:58 ; Search time 231.42 Seconds
(without alignments)
186.377 Million cell updates/sec

Title: US-09-603-713-31
Perfect score: 1708
Sequence: 1 VDEQPLENLDMEYFCTIGI.....ROYFTVDRANNOVLAPVA 326

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_unclassified:*
13: sp_vertebrate:*
14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1558	91.2	387	6 Q9N2D4	Q9n2d4 callithrix
2	1510	88.4	385	6 Q29080	Q29080 sus scrofa
3	1507	88.2	386	6 Q9GMV6	Q9gm6 canis faml
4	1481.5	86.7	390	6 Q9GK10	Q9gk10 camelus dro
5	1465	85.8	386	6 Q9GMV7	Q9gm7 rhinolophus
6	1458	85.4	387	6 Q9GMV8	Q9gm8 sorex ungui
7	1457	85.3	387	6 Q9GMV9	Q9gm9 suncus muri
8	1311	76.8	384	13 Q9DEC2	Q9dec2 xenopus lae
9	1304	76.3	385	13 Q9DEC4	Q9dec4 rana catesb
10	1106	64.8	382	13 Q9PRG9	Q9pr9 gallus gall
11	1074.5	62.9	388	6 Q46524	Q46524 felis silve
12	1041.5	61.0	381	6 Q9N2D2	Q9n2d2 callithrix
13	1025.5	60.0	381	6 Q9GK11	Q9gk11 camelus dro
14	1021.5	59.8	380	6 Q28950	Q28950 sus scrofa
15	1020.5	59.7	387	11 Q9JKE6	Q9jke6 mus musculu
16	1019	59.7	378	13 Q9JKE9	Q9jke9 pleuronecte
17	1013.5	59.3	387	11 Q9JXJ2	Q9jjx2 rattus norv
18	1008	59.0	388	6 Q46523	Q46523 equus zebra
19	998.5	58.5	366	6 Q9N1P5	Q9n1p5 bubalus bub

20	986.5	57.8	379	11 Q9JXJ1	Q9jjx1 rattus norv
21	985.5	57.7	384	13 Q91322	Q91322 rana catesb
22	980	57.4	383	13 Q9DEC3	Q9dec3 xenopus lae
23	973	57.0	376	13 Q9PUR8	Q9pur8 pleuronecte
24	927	54.3	388	6 Q9GMV2	Q9gm2 oryctolagus
25	922.5	54.0	389	6 Q9MYK3	Q9myk3 sus scrofa
26	921.5	54.0	389	6 Q9MYK2	Q9myk2 sus scrofa
27	916	53.6	389	6 Q9MYK5	Q9myk5 suncus muri
28	912	53.4	388	6 Q9N2D3	Q9n2d3 callithrix
29	911	53.3	375	6 Q46500	Q46500 bos taurus
30	908.5	53.2	387	6 Q46496	Q46496 bos taurus
31	908	53.2	389	13 Q9W643	Q9w643 gallus gall
32	908	53.2	389	13 Q9PWK1	Q9pwk1 gallus gall
33	900	52.7	389	6 Q9GMV3	Q9gm3 rhinolophus
34	899	52.6	383	13 Q9DE45	Q9de45 salvelinus
35	899	52.6	387	13 Q9DDV5	Q9ddv5 salvelinus
36	898.5	52.6	345	6 Q27951	Q27951 bos taurus
37	898	52.6	389	6 Q9GMV4	Q9gm4 sorex ungui
38	880	51.5	370	6 Q9TTW1	Q9ttw1 bos taurus
39	877.5	51.4	379	6 Q9TTW6	Q9ttw6 capra hircu
40	876.5	51.3	386	6 Q9TTW8	Q9ttw8 capra hircu
41	875	51.2	381	6 Q9TTV6	Q9ttv6 bos taurus
42	871.5	51.0	380	6 Q9TTV9	Q9ttv9 bos taurus
43	868.5	50.8	380	6 Q9TTV8	Q9ttv8 bos taurus
44	862.5	50.5	376	6 Q9TTX4	Q9ttx4 capra hircu
45	854.5	50.0	397	6 Q28546	Q28546 ovis aries

ALIGNMENTS

RESULT 1

Q9N2D4 PRELIMINARY; PRT: 387 AA.
AC Q9N2D4: 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE PEPsinogen A (EC 3.4.23.1).
OS Callithrix jacchus (Common marmoset).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae; Callithrix.
OX NCBI_TaxID=9483;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=GASTRIC MUCOSA;
RA Kageyama T.;
RT "New World monkey pepsinogens A and C, and prochymosins. Purification,
RT characterization of enzymatic properties, cDNA cloning, and molecular
RT evolution.";
RL J. Biochem. 0:0-0(2000).
CC -!- SIMILARITY: THE PROTEASE BELONGS TO THE PEPTIDASE FAMILY A1; ALSO
CC KNOWN AS THE EUKARYOTIC ASPARTYL PROTEASES FAMILY.
DR EMBL; AB038384; BAA90871.1;
DR InterPro; IPR001461;
DR InterPro; IPR001969;
DR Pfam; PF00026; asp; 1.
DR PRINTS; PR00792; PEPsin.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
KW Aspartyl protease; Hydrolase.
SQ SEQUENCE 387 AA; 41563 MW; 7A7968AA568464BD CRC64;

Query Match 91.2%; Score 1558; DB 6; Length 387;

Best Local Similarity 89.3%; Pred. No. 1.8e-104;

Matches 291; Conservative 20; Mismatches 13; Indels 0; Gaps 0;

Qy 1 VDEQPLENLDMEYFCTIGTIGTPAQDFTVFTDFTGSSNLWVPVSVYCSLACTNHNRFNPD 60

Db 62 IANQPLVNLDMYEGTIGTIGTPAQDFTVFTDFTGSSNLWVPVSVYCSLACTNHNRFNPD 121

Qy 61 SSTYSTSETVTSVITGTSMTGILGYDVTQVQGGISDTNQIFGLSETPGSLFYAPFDGI 120

Db	122	SSTYQATSTQSLT	AIYATGSGW	TGILGYDVT	VQVGGIADT	NTQIFGLST	EPGSGFLYYSPFDGI	181
Qy	121	LGLAYPSIS	SSSGATPV	FDNIWNQ	GLVSQDL	FSVYLSAD	QQSGSVVIFGGIDSSYYTGSLN	180
Db	182	LGLAYPSIS	SSSGATPV	FDNIWNQ	GLVSQDL	FSVYLSAD	QQSGSVVIFGGIDSSYYTGSLN	241
Qy	181	WVPVTVEGW	QITVDSIT	WNGSAICA	EGCQAI	VDVTGSL	TGTPSPITANQSDIGASEN	240
Db	242	WVPVSAEGW	QITVDSIT	WNGSAICA	EGCQAI	VDVTGSL	TGTPSPITANQSDIGASEN	301
Qy	241	SDGDMVVS	CSATSSL	PDIVFT	TNGVQY	PPVP	SAYTILOSEGSCISGFCQGNLPTESGELWI	300
Db	302	SNMGVMVVS	CSATSSL	PDIVFT	TNGVQY	PPVP	SAYTILOSEGSCISGFCQGNLPTESGELWI	361
Qy	301	LGDVFIRQY	FVFD	RANNOV	GLAPVA	326		
Db	362	LGDVFIRQY	FVFD	RANNOV	GLAPVA	367		
RESULT	2							
ID	Q29080	PRELIMINARY;	PRT;	385	AA.			
AC	Q29080;							
DT	01-NOV-1996	(tREMBLrel. 01, Created)						
DT	01-NOV-1996	(tREMBLrel. 01, Last sequence update)						
DT	01-MAR-2001	(tREMBLrel. 16, Last annotation update)						
DE	PEPSINOGEN A	PRECURSOR.						
OS	Sus scrofa	(Pig).						
OC	Eukaryota;	Metazoa;	Chordata;	Craniata;	Vertebrata;	Euteleostomi;		
OC	Mammalia;	Eutheria;	Cetartiodactyla;	Sulna;	Suidae;	Sus.		
NCBI_Taxid=	9823;							
ON	[1]							
RP	SEQUENCE FROM N.A.							
RP	MEDLINE=89174702;	PubMed=2494172;						
RT	Lin X.L., Wong R.N.S., Tang J.;							
RT	"Synthesis, purification, and active site mutagenesis of recombinant							
RT	porcine pepsinogen."							
RL	J. Biol. Chem. 264:4482-4489(1989).							
CC	-1- SIMILARITY: THE PROTEASE BELONGS TO THE PEPTIDASE FAMILY A1; ALSO							
CC	KNOWN AS THE EUKARYOTIC ASPARTYL PROTEASES FAMILY.							
DR	EMBL; J04601; AAA31096.1; -							
DR	HSSP; P00791; 1PSA.							
DR	InterPro; IPR001461; -							
DR	InterPro; IPR001791; -							
DR	InterPro; IPR001969; -							
DR	Pfam; PF00026; asp; 1.							
DR	PROSITE; P00792; PEPsin.							
DR	PROSITE; PS00141; ASP_PROTEASE; 2.							
KW	Aspartyl protease; Hydrolase; Signal.							
FT	SIGNAL	1	35	POTENTIAL.				
FT	CHAIN	16	385	POTENTIAL.				
FT	CHAIN	60	385	POTENTIAL.				
SQ	SEQUENCE	385	AA;	41310	MM;	9A8C406AD5877708	CRC64;	

OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.
OX	NCBI_taxid=9838;
OX	[1]
RP	SEQUENCE FROM N.A.
RP	TISSUP-STOMACH MUCOSA;
RC	Kappeler S.R., Farah Z., Puhan Z.;
RA	"Camel (Camelus dromedarius) Chymosin and Pepsin as Renneting Enzymes for Camel Milk.";
RT	Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RL	EMBL: AJ131678; CAC19555.1; -
RL	Signal; Hydrolase.
KW	Signal; Hydrolase.
KW	Signal; Hydrolase.
FT	SIGNAL 1 63 POTENTIAL.
FT	CHAIN 64 390 PEPSIN A.
FT	SEQUENCE 390 AA; 42117 MW; 7A679DDF946D7F8D CRC64:

Query Match	86.7%	Score 1481.5;	DB 6;	Length 390;
Best Local Similarity	85.6%	Pred. No. 5.9e-99;		
Matches 279; Conservative	26;	Mismatches 20;	Indels 1;	Gaps 1;

Qy	1	VDEQPLENYLDMEYFGTIGTGPQAQDFTVVFDTGSSNLWVPVYCSSLACTNHNRENPE	60
Db	64	LDEQPLENYLDTEYFGTISGTTPAQNTWIFDGTSSNLWVPVSIYCSSSACTNHNRENPEE	123
Qy	61	SSYQSTSETVSITYCTGSMTGILGYDVTQVGGISDTNQIFGLSETEPGSFLYYAPFDGI	120
Db	124	SSYQGTDEPLSYITYCTGSMTGILGYDVTQVGGISDVNQIFGLSETEPGSFLYYAPFDGI	183
Qy	121	LGLAYPSISGGATPVFDNIWNQGLVSQDLFSVYLSADDOGSVVFPGIDSSYYTGSIN	180
Db	184	LGLAYPSISGGCTPVDNIWDEGLISEDLFSVYLSNDESGSVVFPGIDSSYYTGSIN	243
Qy	181	WVPVTEGYWQITVDSITMNGEAIACAECQAIQVDTGSLTGPSTPANTQSDIYGAEN	240
Db	244	WVPVSEGYWQITVDSITWEGESIACSCCQAIQVDTGSLTGPSTPANTQSDIYGAEN	303
Qy	241	SGDMVWSCAISLSPDVIYFTINGVQYVPPPSAYILOEBSGCSIGFQGMNLPTESGELWI	300
Db	304	SYGDMVWSCSISSLNIVFTINGVQYPLSPSAYILESDDCSTCFEGMDL-SSSEELWI	362
Qy	301	LGDDVFIRQYFTVFDNRANNOVGLAPVA	326
Db	363	LGDDVFIRQYFTVFDNRANNOVGLAAVA	388

RESULT	5	
Q9GMV7		
ID	PRELIMINARY;	PRT; 386 AA.
AC	Q9GMV7;	
DT	01-NAR-2001 (TREMBLrel. 16, Created)	
DT	01-NAR-2001 (TREMBLrel. 16, Last sequence update)	
DT	01-NAR-2001 (TREMBLrel. 16, Last annotation update)	
DE	PEPSINOGEN A.	
GN	PGNA.	
OS	Rhinolophus ferrumequinum.	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Chiroptera; Microchiroptera; Rhinolophidae;	
OC	Rhinolophus.	
OX	NCBI_TaxID=59479;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	Narita Y., Oda S., Takenaka O., Kageyama T.;	
RT	*Phylogenetic position of Insectivora inferred from the cdna sequences	
RT	of pepsinogen A and C.;	
RL	Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AB047245; BAB11751.1; -;	
RD	SEQUENCE 386 AA; 41591 MW; 917EE04D3166C3A4 CRC64;	
SQ		

Query Match 85.8%; Score 1465; DB 6; Length 386;
Best Local Similarity 83.0%; Pred. No. 8.9e-98;
Matches 268; Conservative 35; Mismatches 20; Indels 0; Gaps 0;

Qy	4	QPLENYLDMEYFGTIGTCTPAQDFVTVVDFDGTSSNLWVPVSYVCSSLACTNHNRRNPEDSST	63
Db	64	QPLENYLDMEYFGTIGTCTPAQDFVTVVDFDGTSSNLWVPVSYVCSSLACTNHNRRNPEDSST	123
Qy	64	YQSTSETVSIPTYGTSMTGILGYDVTQVGGISDTNOIFGLSETPGCSFLYYAPFDGILGL	123
Db	124	YQGTNQLSVAYGTGTSMTGILGYDVTQVGGITDTNOIFGLSETPGCSFLYYAPFDGILGL	183
Qy	124	AYPSISSGATPVEDNTWNOGLVVSQDLFSVYLSADQSGSVWTFGGIDSSYTGSLNWVP	183
Db	184	AYPSIASSGATPVEDNTWNOGLVVSQDLFSVYLSADQSGSVWTFGGIDSSYTGSLNWVP	243
Qy	184	VTBEGYIWOITVDSITTMNCEAIAACEGCOAIVDTGTSLLTPTSPIANIQSDIGASENSDG	243
Db	244	LSSETYIWOITVDSITTMNQVIACSGSCOAIVDTGTSLLSPTNAIASIQYIGASQANAG	303
Qy	244	DMVYVCSAISLPDIVETINGVQVPPPSAYILQSEGCSCISGQGNLPTSESELWILGD	303
Db	304	EMVYVCSAINFLPNIVETINGVQVPLPPSAIVLQSQOQCTSGFGQMDIPTSSGELWILGD	363
Qy	304	VFIQYITVDFDRANNQVGLAPVA	326
Db	364	VFIQYITVDFDRNNQVGLAPVA	386

```

RESULT      6
QG9MW8
ID          Q9GCMY8
AC          PRELIMINARY;          PRT;      387 AA.
DT          Q9GCMY8;
DT          01-MAR-2001 (T-EMBLrel. 16, Created)
DT          01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT          01-MAR-2001 (T-EMBLrel. 16, Last annotation update)
DE          PEPSINOGEN A.
GN          PGNA.
OS          Sorex unguiculatus (Long-clawed shrew) .
OC          Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;
OC          Mammalia; Eutheria; Insectivora; Soricidae; Sorex.
OX          NCBI_TaxID=62275;
RX          [1]
RP          SEQUENCE FROM N.A.
RA          Narita Y., Oda S., Takenaka O., Kageyama T.;
RT          "Phylogenetic position of insectivora inferred from the cDNA sequences
RT          of pepsinogen A and C.";
RL          Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
RL          EMBL; AB047244; BAB11750.1; -.
SQ          SEQUENCE      387 AA;  41514 MW;  F2EB2E331FAA24BF  CRC64;

```

Query Match	85.4%	Score 1458;	DB 6;	Length 387;
Best Local Similarity	83.0%;	Pred. No. 2.9e-97;		
Matches 268; Conservative	33;	Mismatches 22;	Indels 0;	Gaps 0;

QY	4	QPLNLYDM	EYF	YGT	IGT	GT	PAO	D	T	V	V	F	D	T	G	S	N	L	W	P	S	V	S	S	L	A	C	T	N	H	N	R	F	N	P	E	D	S	S	T	63													
Db	65	QPLVNYDM	EYF	YGT	IS	T	P	P	O	E	F	I	F	D	T	G	S	N	L	W	P	S	I	C	S	S	P	A	C	S	N	H	N	R	F	D	P	O	K	S	T	124												
QY	64	YQSTSE	T	S	T	I	T	Y	G	T	G	S	M	T	G	I	L	G	V	D	T	V	Q	V	G	G	I	S	D	T	N	Q	I	F	G	L	S	E	T	E	P	G	S	F	L	Y	A	P	F	D	G	I	L	123
Db	125	FKPTSC	V	S	I	A	T	G	S	M	T	G	V	L	G	Y	D	T	V	Q	V	A	G	I	A	D	T	N	Q	I	F	G	L	S	O	S	E	P	G	S	F	L	Y	S	P	F	D	G	I	L	184			
QY	124	AYPSIS	S	S	G	A	T	P	V	D	N	I	N	Q	L	V	S	O	D	L	F	S	V	L	S	A	D	D	O	G	S	V	V	I	F	G	I	D	S	S	Y	T	G	S	L	N	W	P	183					
Db	185	AYPSIS	S	S	G	A	T	P	V	D	N	I	N	Q	L	V	S	O	D	L	F	S	V	L	S	S	N	D	O	G	S	V	V	I	F	G	I	D	S	S	Y	T	G	S	L	N	W	P	244					
QY	184	VTEG	Y	W	Q	I	T	V	D	S	I	T	M	G	E	A	I	A	C	E	C	O	A	I	V	D	T	G	S	P	T	I	A	N	T	O	S	I	D	I	G	A	S	E	N	S	D	G	243					
Db	245	LSSEG	Y	W	Q	I	T	V	D	S	I	T	M	G	O	S	I	A	C	E	C	O	A	I	V	D	T	G	S	L	S	G	P	T	A	N	T	O	S	I	G	A	S	E	N	S	O	G	304					
QY	244	DMV	S	C	A	T	S	S	L	P	D	I	V	T	T	I	N	G	V	P	P	S	A	Y	I	L	O	S	E	G	S	C	I	S	G	F	O	G	M	N	I	P	T	E	S	G	E	L	W	I	L	D	303	
Db	305	QMAV	S	C	S	I	K	N	L	P	D	I	V	T	T	I	N	G	V	P	A	S	A	Y	I	L	O	S	E	G	S	C	I	S	G	F	O	G	M	N	I	P	T	E	S	G	E	L	W	I	L	D	364	

```
RT 304 VFIRQYFTVDRANNOVGLAPVA 326
RT 365 VFIRQYFTVDRANNOVGLAPVA 387
DB

RESULT 7
Q9GMY9 PRELIMINARY; PRT; 387 AA.
AC
AC Q9GMY9;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE PEPSINOGEN A.
GN PGNA.
OS Suncus murinus (House shrew) (Musk shrew).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Insectivora; Soricidae; Suncus.
OX NCBI_TaxID=9378;
RN
RP SEQUENCE FROM N.A.
RA Narita Y., Oda S., Takenaka O., Kageyama T.;
RT "Phylogenetic position of Insectivora inferred from the cDNA sequences
of pepsinogen A and C.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB047243; BAB11749.1; -. 7F9FF818DD8541CE CRC64;
SQ SEQUENCE 387 AA; 41576 MW; 7F9FF818DD8541CE CRC64;

Query Match 85.3%; Score 1457; DB 6; Length 387;
Best Local Similarity 82.1%; Pred. No. 3.4e-97;
Matches 266; Conservative 36; Mismatches 22; Indels 0; Gaps 0;

QY 3 EOPLENYDMEYFGTIGTTPAQDFTVVDGSSNLWVPVYVCSLACTNHNRPEDSS 62
DB 64 DQPLVNDMEYFGTIGTTPAQDFTVVDGSSNLWVPVYVCSLACTNHNRPEDSS 123

QY 63 TVQSTSETVITYGTGSMGILGYDTVOVGGISDTNQIFGLSETPGSFLYAPFDGILG 122
DB 124 TFQSTSETVITYGTGSMGILGYDTVOVGGISDTNQIFGLSETPGSFLYAPFDGILG 183

QY 123 LAYPSISSGATPVFDNIWNOGLVSDLFVSVLSDGSGVIFGGIDSSYITGNLWV 182
DB 184 LAYPNASSGATPVFDNIWNOGLVSDLFVSVLSDGSGVIFGGIDSSYITGNLWV 243

QY 183 PVTVEGWQITVDSITMNGEATACAGCAIYDVTGTSLLTGTPTSIANTQSDIGASENSD 242
DB 244 PLSEGGWQITVDSITMNGEATACAGCAIYDVTGTSLLTGTPTSIANTQSDIGASENSD 303

QY 243 GDMVSCSAISSLPDIVFTINGVQVPPPSAYILQSEGSCISGFGQMNLPTEGSELWILG 302
DB 304 GDMVSCSAISSLPDIVFTINGVQVPPPSAYILQSEGSCISGFGQMNLPTEGSELWILG 363

QY 303 DVFIQYFTVDRANNOVGLAPVA 326
DB 364 DVFIQYFTVDRANNOVGLAPVA 387

RESULT 8
Q9DEC2 PRELIMINARY; PRT; 384 AA.
AC
AC Q9DEC2;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE PEPSINOGEN A.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN
RP SEQUENCE FROM N.A.
RA Ikuzawa M., Inokuchi T., Kobayashi K., Yasumasu S.;
RT "Molecular Cloning of pepsinogens in Adult Xenopus laevis and Bullfrog
Rana catesbeiana.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB045376; BAB20092.1; -. 5DC8914FC9CEC603 CRC64;
SQ SEQUENCE 385 AA; 41702 MW; 5DC8914FC9CEC603 CRC64;
```

```
RT "Molecular Cloning of Pepsinogens in Adult Xenopus laevis and Bullfrog
Rana catesbeiana.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB045380; BAB20798.1; -. 729DD2E7953D9072 CRC64;
SQ SEQUENCE 384 AA; 41888 MW; 729DD2E7953D9072 CRC64;

Query Match 76.8%; Score 1311; DB 13; Length 384;
Best Local Similarity 71.8%; Pred. No. 1.1e-86;
Matches 232; Conservative 57; Mismatches 34; Indels 0; Gaps 0;

QY 4 QPLENYDMEYFGTIGTTPAQDFTVVDGSSNLWVPVYVCSLACTNHNRPEDSS 63
DB 62 ETLQNYMDIEYGTISIGTTPQDFTVVDGSSNLWVPVYVCSLACTNHNRPEDSS 121

QY 64 YQSTSETVITYGTGSMGILGYDTVOVGGISDTNQIFGLSETPGSFLYAPFDGILG 123
DB 122 FQATNTPVSIQYGTGSMGILGYDTVOVGGISDTNQIFGLSETPGSFLYAPFDGILG 181

QY 124 AVPSISSGATPVFDNIWNOGLVSDLFVSVLSDGSGVIFGGIDSSYITGNLWVP 183
DB 182 AFPSIASSGATPVFDNIWNOGLVSDLFVSVLSDGSGVIFGGIDSSYITGNLWVP 241

QY 184 VTVEGWQITVDSITMNGEATACAGCAIYDVTGTSLLTGTPTSIANTQSDIGASENSD 243
DB 242 LTAETWQITVDSITMNGEATACAGCAIYDVTGTSLLTGTPTSIANTQSDIGASENSD 301

QY 244 DMVSCSAISSLPDIVFTINGVQVPPPSAYILQSEGSCISGFGQMNLPTEGSELWILG 303
DB 302 QYVINCNNISNMTVFTINGVQVPLSPSAYVRQNOQCGSGFQAMNLPNTSGDLWILG 361

QY 304 VFIRQYFTVDRANNOVGLAPVA 326
DB 362 VFIRQYFTVDRANNOVGLAPVA 384

RESULT 9
Q9DEC4 PRELIMINARY; PRT; 385 AA.
AC
AC Q9DEC4;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE PEPSINOGEN A.
OS Rana catesbeiana (Bull frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Ranioidea; Rana;
OX NCBI_TaxID=8400;
RN
RP SEQUENCE FROM N.A.
RA Ikuzawa M., Inokuchi T., Kobayashi K., Yasumasu S.;
RT "Molecular Cloning of pepsinogens in Adult Xenopus laevis and Bullfrog
Rana catesbeiana.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB045376; BAB20092.1; -. 5DC8914FC9CEC603 CRC64;
SQ SEQUENCE 385 AA; 41702 MW; 5DC8914FC9CEC603 CRC64;
```

```
Query Match 76.3%; Score 1304; DB 13; Length 385;
Best Local Similarity 72.4%; Pred. No. 3.4e-86;
Matches 234; Conservative 48; Mismatches 41; Indels 0; Gaps 0;

QY 4 QPLENYDMEYFGTIGTTPAQDFTVVDGSSNLWVPVYVCSLACTNHNRPEDSS 63
DB 63 EPLQNYMDIEYGTISIGTTPQDFTVVDGSSNLWVPVYVCSLACTNHNRPEDSS 122

QY 64 YQSTSETVITYGTGSMGILGYDTVOVGGISDTNQIFGLSETPGSFLYAPFDGILG 123
DB 123 FQATNTPVSIQYGTGSMGILGYDTVOVGGISDTNQIFGLSETPGSFLYAPFDGILG 182

QY 124 AVPSISSGATPVFDNIWNOGLVSDLFVSVLSDGSGVIFGGIDSSYITGNLWVP 183
DB 183 AFPSIASSGATPVFDNIWNOGLVSDLFVSVLSDGSGVIFGGIDSSYITGNLWVP 242
```

Qy	184	VTVEGYWQIVTDSITMNGEAIACAECQAIVDTGTSLLTGPTSPIANQSDIGASENSDG	24
Db	243	LTAETVWQIVTDSISIGGOVYACSGSCSAIVDTGTSLLAGSPPIANQYICGNQDSMG	302
Qy	244	DMYVSCSAISLPLDITFTINGVQVPPPSAYILQSGSCISGFGQGNLPTSEGELWILGD	303
Db	303	QYVINCNNISNMPTVFTINGVQVPLPASAYVRQSQSGTSGFQANMLPTSSGDLWILGD	362
Qy	304	VFIQYFTVFDNRANNVOGLAPVA	326
Db	363	VFIREYVVFDRANNYAMAPVA	385
RESULT 10			
Q9PRG9. PRELIMINARY; PRT; 382 AA.			
ID	Q9PRG9		
AC	Q9PRG9		
DT	01-MAY-2000 (T-EMBLrel. 13, Created)		
DT	01-MAY-2000 (T-EMBLrel. 13, Last sequence update)		
DT	01-MAR-2001 (T-EMBLrel. 16, Last annotation update)		
DE	PEPSTNOGEN A.		
GN	CPGA.		
OS	Gallus gallus (Chicken).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;		
OC	Gallus.		
OX	NCBI_TaxID=9031;		
RP	SEQUENCE FROM N.A.		
RP	Sakamoto N., Yasugi S.;		
RA	Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.		
RL	[2]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=98440813; PubMed=9753645;		
RT	Sakamoto N., Saiga H., Yasugi S.;		
RT	"Analysis of temporal expression pattern and cis-regulatory sequences		
RT	of chicken pepsinogen A and C.;"		
RT	Biochem. Biophys. Res. Commun. 250:420-424(1998).		
CC	-1- SIMILARITY: THE PROTEASE BELONGS TO THE PEPTIDASE FAMILY A1; ALSO		
CC	KNOWN AS THE EUKARYOTIC ASPARTYL PROTEASES FAMILY.		
DR	EMBL; AB025283; BAA77268.1; -;		
DR	EMBL; AB025281; BAA76891.1; -;		
DR	HSSP; P00794; 4CMS.		
DR	InterPro; IPR001461; -;		
DR	InterPro; IPR001969; -;		
DR	Pfam; PF00026; asp. 1.		
DR	PRINTS; PR00792; PEP SIN.		
DR	PROSITE; PS00141; ASP_PROTEASE; 2.		
DR	Aspartyl protease; Hydrolase.		
SK	SEQUENCE 382 AA; 42056 MW; C3C7EB8CA4DD3E28 CRC64;		
Query Match 64.8%; Score 1106; DB 13; Length 382;			
Best Local Similarity 63.8%; Pred. No. 5.8e-72;			
Matches 206; Conservative 43; Mismatches 70; Indels 4; Gaps			
Qy	4	QPLENYLDMYFETIGTGTPAQDFTVVFDTGSSNLWVPSVYCSSLACTNHNRFNPEDSSST	63
Db	64	EPMTNMDASYTYTISIGTPQDQDFTVFDGSSNLWVPSYICKSSACSNHRRFDPKSKST	123
Qy	64	YQSTSETVSTYCTGSMTGILGYDVTVOVGISDTHQIFGLSETERGSLFYAPDPGILGL	123
Db	124	YVSTNETVYAYTGSMGILGYDVTVAVSSIDVQIQIFGLSETERGSPFYFCNFGILGL	183
Qy	124	AYPSISSSGATPVFDINWQVSLQDLFSVYLSADDSQGVVIFGIDSSVYTGSLNWVP	183
Db	184	AFPSISSSGATPVFDNMMSOHLVAQDLFSVYLSKDGETGSFVIFGIDPNYTKGIWVP	243
Qy	184	VTVEGYWQIVTDSITMNGEAIACAECQAIVDTGTSLLTGPTSPIANQSDIGASENSDG	243
Db	244	LSAETVWQIVTDSITMNGEAIACAECQAIVDTGTSLLTGPTSPIANQSDIGASENSDG	301

Qy	244	DMVYSCSAISSLPDVI	FTINGVQVPPPSAYILQ	SGSCISGFGQGNMLPTES	GELWTLGD 303
Db	302	E--ISCDISKL	PDVTHINGHAF	TLPSAYVLNEDSCMLGF	ENMGCTPTLGEQWTLGD 359
Qy	304	VFIQYFTV	FDNRANNVGLAPVA	326	
Db	360	VFIERYV	FDNRANKVGLSPLS	382	
RESULT 11					
ID	046524	PRELIMINARY;		PRT;	388 AA.
AC	046524;				
DT	01-JUN-1998	(TReMBLrel. 06, Created)			
DT	01-JUN-1998	(TReMBLrel. 06, Last sequence update)			
DT	01-MAR-2001	(TReMBLrel. 16, Last annotation update)			
DE	PREGNANCY-ASSOCIATED GLYCOPROTEIN.				
DE	Felis silvestris catus (Cat).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.				
OX	NCBI_TaxID=9685;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Gan X., Xie S., Green J., Roberts R.M.;				
RL	Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.				
CC	-1- SIMILARITY: THE PROTEASE BELONGS TO THE PEPTIDASE FAMILY A1; ALSO KNOWN AS THE EUKARYOTIC ASPARTYL PROTEASES FAMILY.				
CC	EMBL; AF036953; AB91422.1; -.				
DR	HSP; P00797; ZREN				
DR	InterPro; IPR001461; -.				
DR	InterPro; IPR001969; -.				
DR	Pfam; PF00026; asp. 1.				
DR	PRINTS; PR00792; PEPsin.				
DR	PROSITE; PS00141; ASP.PROTEASE; 2.				
KW	Aspartyl protease; Hydrolase.				
SQ	SEQUENCE 388 AA; 43030 MW; 4DEA6D7590F9A92C CRC64;				
Query Match 62.9%; Score 1074.5; DB 6; Length 388;					
Best Local Similarity 60.6%; Pred. No. 1.1e-69;					
Matches 198; Conservative 49; Mismatches 67; Indels 13; Gaps					
Qy	4	OPLNYLDM	EYFGTIGT	TPAQDFTVV	FDTGSSNLVSPVSYCSSLACTNHNRFNPDSST 63
Db	65	EPMRNLDL	AYVITIGT	TPQEFKVFID	TGSSDLVWPVSIYCSPACAHNVFNPLRSST 124
Qy	64	YQSTSE	TVSYTYTG	SGMTGILG	YDVTQVGGISDTNIFGLSTPECPGSLFYAPFDGILGL 123
Db	125	FRISGR	PIHLYQ	SGTMSGLAYD	TVRFGLVDVAAGLSLRKPKFWEYAVFDGILGL 184
Qy	124	AYPSISSG	ATPVFD	TINWOGV	QSQDLFSVYLADDSGSGVIFGGIDSYTGSNLWNPV 183
Db	185	AYPSLSUR	GVPPVFD	NLWKGIL	QSLELFAYLKSKEGSGVWVFGVDHSHYSGDLNWNVP 244
Qy	184	VTVEGYW	IOITVD	STWN	GEAIAECGCOATVDGTGTSLLTGPTSPATNIQSDIGASENSDG 243
Db	245	VSKRLY	WQLSND	SISMN	GEVIACDGCQALIDTGTSELLGPHSVFNIQIIGANQISYG 304
Qy	244	DMVYSCSAISSLPD	VI	FTINGVQVPPPSAYILQ	S-EGSCISGFGQGNMLPTESG----- 296
Db	305	EYVVDCA	DAAN	TLDPDVI	FTINGIDYPPVPASAYIQEGPQGTCTYSGFD-----ESGDSLLVS 358
Qy	297	ELWILGD	VETROY	FTV	FDNRANNOVGLA 323
Db	359	DSWILGD	VFLRLY	FTV	FDRENNRIGLA 385
RESULT 12					
ID	Q9N2D2	PRELIMINARY;		PRT;	381 AA.
AC	Q9N2D2;				
DT	01-OCT-2000	(TReMBLrel. 15, Created)			
DT	01-OCT-2000	(TReMBLrel. 15, Last sequence update)			

DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE PROCHYMOSIN (EC 3.4.23.4).
 OS Callithrix jacchus (Common marmoset).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae; Callithrix.
 OX NCBI_TaxID=9483;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=GASTRIC MUCOSA;
 RA Kageyama T.;
 RT "New World monkey pepsinogens A and C, and prochymosins. Purification,
 RT characterization of enzymatic properties, cDNA cloning, and molecular
 RT evolution.";
 RL J. Biochem. 0:0-0(2000).
 CC -|- KNOWN AS THE EUKARYOTIC ASPARTYL PROTEASES FAMILY.
 DR EMBL: AB038386; BA90873.1; -;
 DR InterPro; IPR001461; -;
 DR InterPro; IPR001969; -;
 DR Pfam; PF00026; asp; 1.
 DR PRINTS; PR00792; PEPsin.
 DR PROSITE; PS00141; ASP_PROTEASE; 2.
 KW Aspartyl protease; Hydrolase.
 SQ SEQUENCE 381 AA; 41896 MW; C5820C74C97BB96B CRC64;

Query Match 61.0%; Score 1041.5; DB 6; Length 381;
 Best Local Similarity 61.0%; Pred. No. 2.5e-67;
 Matches 194; Conservative 51; Mismatches 68; Indels 5; Gaps 2;
 QY 6 LENYLDMEVFGTIGTTPAQDTFVVDFTGSSNLWPSVYCSLACTNHNRPEDSSTYQ 65
 DB 66 LTNLDQYFGKIYIGTPPEETVVDFTGSSDLWPSVYCSNACQNHREDPFSKSTFQ 125
 QY 66 STSETVSTYTGSGMTGILGYDTVQVGGISDTNQIFGLSETEPGSFLYAPFDGILGLAY 125
 DB 126 NMDKLSIQYGTGSMQGLGYDTVTVSSIVDPHQTVGLSTQEPGDVFTYSEFDGILGLAY 185
 QY 126 PSTSSSGATPVDNTWNOGLVSQDLFSVLSADDSGSSVIFGGIDSSYYTGSLSNVPVT 185
 DB 186 PSLASEYSPVFDNMMDRLHVAQDLFSVYMSRNEQ-GSMLTGLAIDPSYTTGSLHWIPVT 244
 QY 186 VEGYWOITVDSITMNGEATACAGCOAIVDTGTSLLTGPTSPIANTQSDIGASENSDGM 245
 DB 245 VQIYQFTVDSVTDGVVACDGGCAILLDTGTSMLVGPDSIFNLQQAIGATEQGYGEF 304
 QY 246 VVSCSAISLDPDIFTINGVQVPPSPAYILOSEGSCISGFGOMNLPTESGELWILGDVF 305
 DB 305 DIDCGTSSMPTVVFPEINGKYPPLPSAVTNDQDGFCTSGFQG---DDSSQQWILGDVF 360
 QY 306 IROYFTVFDNRANNOVGLA 323
 DB 361 IREYYSVFDNRANNOVGLA 378

RESULT 13
 Q9GK11 ID Q9GK11 PRELIMINARY; PRT; 381 AA.
 AC Q9GK11;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE CHYMOSIN PRECURSOR (EC 3.4.23.4).
 GN CHYMOSIN.
 OS Camelus dromedarius (Dromedary) (Arabian camel).
 OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.
 OX NCBI_TaxID=9838;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=STOMACH MUCOSA;
 RA Kappeler S.R., Farah Z., Puhon Z.;
 RT "Camel (Camelus dromedarius) Chymosin and Pepsin as Renneting Enzymes

RT for Camel Milk.";
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ131677; CAC19554.1; -;
 KW Signal; Hydrolase.
 FT SIGNAL 1 58 POTENTIAL.
 FT CHAIN 59 381 CHYMOSIN.
 SQ SEQUENCE 381 AA; 42082 MW; 24BADB57B2E7FDD7 CRC64;
 Query Match 60.0%; Score 1025.5; DB 6; Length 381;
 Best Local Similarity 59.8%; Pred. No. 3.6e-66;
 Matches 193; Conservative 50; Mismatches 75; Indels 5; Gaps 2;
 QY 1 VDOQPLENYLDMEVFGTIGTTPAQDTFVVDFTGSSNLWPSVYCSLACTNHNRPED 60
 DB 61 VAREPLTSLDYSQYFGKIYIGTPPEETVVDFTGSSDLWPSVYCKSNVCKNHHREDPDK 120
 QY 61 SSTYQSTSETVSTYTGSGMTGILGYDTVQVGGISDTNQIFGLSETEPGSFLYAPFDGI 120
 DB 121 SSTFRNLGKPLSIHYGTGSMEGFLGYDTVTVSNIVDPNQTVLSTEQPGEVFTYSEFDGI 180
 QY 121 LGLAYPSISSSGATPVDNTWNOGLVSQDLFSVLSADDSGSSVIFGGIDSSYYTGSLSN 180
 DB 181 LGLAYPSLASEYSPVFDNMMDRLHVAQDLFSVYMDRNGQ-GSMLTGLAIDPSYTTGSLH 239
 QY 181 WVPVTVEGYWOITVDSITMNGEATACAGCOAIVDTGTSLLTGPTSPIANTQSDIGASEN 240
 DB 240 WVPVTLQYQWFTVDSVTINGVAVACVGGCAILLDTGTSVLFPGSPSILKIQMAIGATEN 299
 QY 241 SDGMVYSCSAISLDPDIFTINGVQVPPSPAYILOSEGSCISGFGOMNLPTESGELWI 300
 DB 300 RYGEFVDCNGLRSMPTVVFPEINGRDPYPLSPSAYTSKDGQFCTSGFQGDN---NSELWI 355
 QY 301 LGDVFIRQYFTVFDNRANNOVGLA 323
 DB 356 LGDVFIREYYSVFDNRANNOVGLA 378
 RESULT 14
 Q28950 ID Q28950 PRELIMINARY; PRT; 380 AA.
 AC Q28950;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE PREPROCHYMOSIN PRECURSOR (FRAGMENT).
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=GASTRIC;
 RA Foltmann B., Jensen A.L., Loenblad P., Smidt E., Axelsen N.H.;
 RT "A developmental analysis of the production of chymosin and pepsin in
 RT pigs.";
 RL Comp. Biochem. Physiol. 68B:9-13(1981).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=GASTRIC;
 RX MEDLINE=96252892; PubMed=8673731;
 RA Houten G., Madsen M.T., Harlow K.W., Loenblad P., Foltmann B.;
 RT "The primary structure and enzymic properties of porcine prochymosin
 RT and chymosin.";
 RL Int. J. Biochem. Cell Biol. 28:667-675(1996).
 DR EMBL: U14406; AAB08492.1; -;
 DR HSP; P00794; 4CMS.
 DR InterPro; IPR001461; -;
 DR InterPro; IPR001969; -;
 DR Pfam; PF00026; asp; 1.
 DR PRINTS; PR00792; PEPsin.
 DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_2.
 KW Signal.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:51:13 ; Search time 72.75 Seconds
(without alignments)
153.502 Million cell updates/sec

Title: us-09-603-713-31

Perfect score: 1708

Sequence: 1 VDEQPLENLDMEYFCTIGI.....ROYFTVDRANNQGLAPVA 326

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1708	100.0	388	1	PEPA_HUMAN
2	1672	97.9	388	1	PEP2_MACFU
3	1648	96.5	388	1	PEP1_MACFU
4	1646	96.4	388	1	PEPA_MACHU
5	1595	93.4	388	1	PEP4_MACFU
6	1508.5	88.3	386	1	PEPA_PIG
7	1453	85.1	387	1	PEP3_RABIT
8	1382	80.9	387	1	PEP4_RABIT
9	1374	80.4	387	1	PEP2_RABIT
10	1348	78.9	387	1	PEP1_RABIT
11	1102	64.5	367	1	PEPA_CHICK
12	1068.5	62.6	383	1	PEPE_CHICK
13	1053.5	61.7	324	1	PEP1_GADMO
14	1030	60.3	388	1	PEPF_RABIT
15	1021	59.8	388	1	PAG_HORSE
16	1018.5	59.6	381	1	CHYM_SHEEP
17	1008.5	59.0	381	1	CHYM_BOVIN
18	988.5	57.9	396	1	CATE_HUMAN
19	972.5	56.9	398	1	CATE_RAT
20	965.5	56.5	397	1	CATE_MOUSE
21	957.5	56.1	396	1	CATE_RABIT
22	930	54.4	391	1	CATE_CAVPO
23	921	53.9	394	1	PEPC_CAVPO
24	916	53.6	392	1	PEPC_RAT
25	915	53.6	377	1	PEPC_MACFU
26	910	53.3	388	1	PEPC_HUMAN
27	900	52.7	420	1	PAG2_PIG
28	876	51.3	376	1	PAG2_BOVIN
29	844.5	49.4	382	1	PAG1_SHEEP
30	841.5	49.3	389	1	PAG1_PIG
31	825.5	48.3	380	1	PAG1_BOVIN
32	822.5	48.2	412	1	CATD_HUMAN
33	820	48.0	398	1	CATD_CHICK

RESULT 1

PEPA_HUMAN					
ID	PEPA_HUMAN	STANDARD;	PRT;	388	AA.
AC	P00790;				
DT	21-JUL-1986	(Rel. 01, Created)			
DT	21-JUL-1986	(Rel. 01, Last sequence update)			
DT	01-OCT-2000	(Rel. 40, Last annotation update)			
DE	PEPSIN A PRECURSOR (EC 3.4.23.1).				
GN	PGA3 AND PGA4 AND PGA5.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=83161158; PubMed=6300126;				
RA	Sogawa K., Fujii-Kuriyama Y., Mizukami Y., Ichihara Y., Takahashi K.;				
RT	"Primary structure of human pepsinogen gene.";				
RL	J. Biol. Chem. 258:5306-5311(1983).				
RN	[2]				
RP	SEQUENCE FROM N.A. (ISOZYME 5).				
RC	TISSUE=Placenta;				
RX	MEDLINE=89233110; PubMed=2714789;				
RA	Evers M.P.J., Zelle B., Bebelman J.P., van Beusechem V., Kraakman L.,				
RA	Hoffer M.J.V., Pronk J.C., Mager W.H., Planta R.J., Eriksson A.W.,				
RT	Frants R.R.;				
RT	"Nucleotide sequence comparison of five human pepsinogen A (PGA)				
RL	genes: evolution of the PGA multigene family.";				
RN	[3]				
RP	PARTIAL SEQUENCE OF 1-28				
RX	MEDLINE=86059312; PubMed=2415509;				
RA	Ichihara Y., Sogawa K., Takahashi K.;				
RT	"Isolation of human, swine, and rat prepepsinogens and calf				
RT	preprochymosin, and determination of the primary structures of their				
RT	NH2-terminal signal sequences.";				
RN	[4]				
RP	SEQUENCE OF 16-100 (ISOZYMES 2; 3; 3A; 4 AND 5).				
RX	MEDLINE=90130402; PubMed=2515193;				
RA	Athauda S.B.P., Tanji M., Kageyama T., Takahashi K.;				
RT	"A comparative study on the NH2-terminal amino acid sequences and				
RT	some other properties of six isozymic forms of human pepsinogens and				
RL	pepsins.";				
RN	[5]				
RP	SEQUENCE OF 16-68 (ISOZYMES 3 AND 5).				
RX	MEDLINE=89055108; PubMed=3197840;				
RA	Foltmann B.;				
RT	"Activation of human pepsinogens.";				
RL	FEBS Lett. 241:69-72(1988).				
RN	[6]				
RP	X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).				
RX	MEDLINE=95392399; PubMed=7663352;				
RA	Fujinaga M., Chernata M.M., Tarasova N.I., Mosimann S.C.,				
RA	James M.N.G.;				

ALIGNMENTS

34	816.5	47.8	410	1	CATD_MOUSE	P18242	mus musculus
35	816	47.8	407	1	CATD_RAT	P24268	rattus norv
36	788.5	46.2	387	1	ASPP_AEDAE	Q03168	aedes aegypt
37	785	46.0	345	1	CATD_PIG	P00795	sus scrofa
38	737	43.1	419	1	KDAP_MOUSE	O09043	mus musculus
39	729	42.7	420	1	NAP1_HUMAN	O96009	homo sapien
40	701	41.0	473	1	CYPL_CYNCA	P40782	cynara card
41	698	40.9	402	1	RENI_RAT	P08424	rattus norv
42	696	40.7	400	1	RENI_SHEEP	P52115	ovis aries
43	693	40.6	508	1	ASPR_HORVU	P42210	hordeum vul
44	692	40.5	509	1	APR1_ORYSA	Q42456	oryza sativ
45	688	40.3	396	1	CARP_NEUCR	Q01294	neurospora

Crystal structure of human pepsin and its complex with pepstatin.;
 Protein Sci. 4:960-972(1995).
 CC -1- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
 INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
 ALSO CLEAVED TO SOME EXTENT.
 CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
 EUKARYOTIC ASPARTYL PROTEASES FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; J00287; AAA98529.1; -
 DR EMBL; J00279; AAA98529.1; JOINED.
 DR EMBL; J00280; AAA98529.1; JOINED.
 DR EMBL; J00281; AAA98529.1; JOINED.
 DR EMBL; J00282; AAA98529.1; JOINED.
 DR EMBL; J00283; AAA98529.1; JOINED.
 DR EMBL; J00284; AAA98529.1; JOINED.
 DR EMBL; J00285; AAA98529.1; JOINED.
 DR EMBL; J00286; AAA98529.1; JOINED.
 DR EMBL; M26032; AAA60061.1; -
 DR EMBL; M26025; AAA60061.1; JOINED.
 DR EMBL; M26026; AAA60061.1; JOINED.
 DR EMBL; M26027; AAA60061.1; JOINED.
 DR EMBL; M26028; AAA60061.1; JOINED.
 DR EMBL; M26029; AAA60061.1; JOINED.
 DR EMBL; M26030; AAA60061.1; JOINED.
 DR EMBL; M26031; AAA60061.1; JOINED.
 DR PIR; A00980; PEHU.
 DR PIR; A30142; A30142.
 DR PIR; B30142; B30142.
 DR PIR; A22434; A22434.
 DR PIR; PX0023; PX0023.
 DR PIR; PX0024; PX0024.
 DR PIR; PX0025; PX0025.
 DR PIR; PX0026; PX0026.
 DR PIR; PX0027; PX0027.
 DR PIR; S02663; S02663.
 DR PIR; S02664; S02664.
 DR PDB; 1PSN; 20-APR-95.
 DR PDB; 1PSO; 20-APR-95.
 DR MEROPS; A01.001; -
 DR MIM; 169700; -
 DR MIM; 169710; -
 DR MIM; 169720; -
 DR MIM; 169730; -
 DR InterPro; IPR001461; -
 DR InterPro; IPR001969; -
 DR Pfam; PF00026; asp. 1.
 DR PRINTS; PR00792; PEPsin.
 DR PROSITE; PS00141; ASP_PROTEASE; 2.
 KW Hydrolase; Aspartyl protease; Digestion; Stomach; Gastric juice;
 KW Zymogen; Signal; Phosphorylation; 3D-structure; Polymorphism.
 FT SIGNAL 1 15
 FT PROPEP 16 62 ACTIVATION PEPTIDE.
 FT CHAIN 63 388 PEPsin A.
 FT MOD_RES 130 130 PHOSPHORYLATION (POTENTIAL).
 FT ACT_SITE 94 94
 FT ACT_SITE 277 277
 FT DISULFID 107 112
 FT DISULFID 268 272
 FT DISULFID 311 344
 FT VARIANT 28 28 L->F (IN ISOZYME 5).
 FT FTid-VAR_006481.
 FT E->K (IN ISOZYMES 3A, 4 AND 5).
 FT FTid-VAR_006482.
 FT VARIANT 92 92 V->L (IN ISOZYMES 4 AND 5).
 FT VARIANT 92 92

FT VARIANT 222 222 /FTid-VAR_006483.
 FT O->K.
 FT FTid-VAR_006484.
 FT A->T.
 FT FTid-VAR_006485.
 FT L->V.
 FT FTid-VAR_006486.
 FT D->E.
 FT FTid-VAR_006487.
 SQ SEQUENCE 388 AA; 41977 MW; C9CB89BA08F4D78B CRC64;
 Query Match 100.0%; Score 1708; DB 1; Length 388;
 Best Local Similarity 100.0%; Pred. No. 4.3e-117;
 Matches 326; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VDEQPLENYLDMYFEGTIGTGAQDFTVVFDTGSSNLWVPSVYCSSLACTNHNRFNPD 60
 DB 63 VDEQPLENYLDMYFEGTIGTGAQDFTVVFDTGSSNLWVPSVYCSSLACTNHNRFNPD 122
 QY 61 SSTVQSTSETVSTYTGSGMTGILGYDVTQVGGISDNTQIFGLSETEPGSLFYVAPDGI 120
 DB 123 SSTVQSTSETVSTYTGSGMTGILGYDVTQVGGISDNTQIFGLSETEPGSLFYVAPDGI 182
 QY 121 LGLAYPSISSSGATPVFDNIWQGLVSQDLFSVYLSADDSQSGSVIFGGIDSSYTTGSLN 180
 DB 183 LGLAYPSISSSGATPVFDNIWQGLVSQDLFSVYLSADDSQSGSVIFGGIDSSYTTGSLN 242
 QY 181 WVPVTVEGYWQITVDSITMNGEAIACAGCOAIVDTGTSLLTGTSPIANIQSDIGASEN 240
 DB 243 WVPVTVEGYWQITVDSITMNGEAIACAGCOAIVDTGTSLLTGTSPIANIQSDIGASEN 302
 QY 241 SDGDMVYSCSAISSLPDIVFTINGVQVPPPSAVILOSGSCISGFGOMNLPTESGELWI 300
 DB 303 SDGDMVYSCSAISSLPDIVFTINGVQVPPPSAVILOSGSCISGFGOMNLPTESGELWI 362
 QY 301 LGDVFIRQYFTVFDNRANNOVGLAPVA 326
 DB 363 LGDVFIRQYFTVFDNRANNOVGLAPVA 388
 RESULT 2
 ID PEP2_MACFU STANDARD; PRT; 388 AA.
 AC P27677;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE PEPsin A-2/A-3 PRECURSOR (EC 3.4.23.1) (PEPSIN III-2/III-1).
 OS Macaca fuscata fuscata (Japanese macaque).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecoidea; Macaca.
 OC NCBI_TaxID=9543;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 16-70.
 RC TISSUE-Gastric mucosa;
 RX MEDLINE=92037645; PubMed=1935977;
 RA Kageyama T., Tanabe K., Koizumi O.;
 RT "Development-dependent expression of isozymes of monkey
 pepsinogens and structural differences between them";
 RL Eur. J. Biochem. 202:205-215(1991).
 CC -1- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
 INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
 ALSO CLEAVED TO SOME EXTENT.
 CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.
 CC -1- DEVELOPMENTAL STAGE: PEP A-2 IS PREDOMINANT AT THE 4-MONTH STAGE.
 CC PEP A-3 IS PREDOMINANT AT FETAL STAGES.
 CC -1- PTM: PEP A-2 IS PHOSPHORYLATED, BUT NOT PEP A-3.
 CC -1- PTM: EACH PEPsinOGEN IS CONVERTED TO CORRESPONDING PEPsin AT PH
 2.0 IN PART AS A RESULT OF THE RELEASE OF A 47 AA ACTIVATION
 SEGMENT AND IN PART AS A RESULT OF STEPWISE PROTEOLYTIC CLEAVAGE
 VIA AN INTERMEDIATE FORM(S).
 CC

CC -1- MISCELLANEOUS: THE EXPRESSION OF PEPsin GENES IS REGULATED BY
CC HORMONES AND RELATED SUBSTANCES.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC -1- CAUTION: IT IS NOT KNOWN IF THIS IS PEP A-2 OR PEP A-3.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X59755; CAA42427.1; --
CC PIR; S16064; S16064.
CC PIR; S19684; S19684.
CC HSSP; P00790; IFSO.
CC MEROPS; A01.001; --
CC InterPro; IPR001461; --
CC InterPro; IPR001969; --
CC Pfam; PF00026; asp; 1.
CC PRINTS; PR00792; PEPsin.
CC PROSITE; PS00141; ASP_PROTEASE; 2.
CC Hydrolase: Aspartyl protease; Digestion; Stomach; Gastric juice;
KW Zymogen; Multigene family; Phosphorylation; Signal.
FT SIGNAL 1 15
FT PROPEP 16 40
FT CHAIN 63 388
FT ACT_SITE 94 94
FT ACT_SITE 277 277
FT DISULFID 107 112
FT DISULFID 268 272
FT DISULFID 311 344
FT MOD_RES 130 130
SQ SEQUENCE 388 AA; 41703 MW; 706F7ED50FF59C0D CRC64;

Query Match 97.9%; Score 1672; DB 1; Length 388;
Best Local Similarity 96.6%; Pred. No. 1.8e-114;
Matches 315; Conservative 10; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VDEQPLENLYMEYGTGIGTGAQDFVFDGTSSNLWVSVYCSSLACTNHNRFNPED 60
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|
63 IDEQPLENLYMEYGTGIGTGAQDFVFDGTSSNLWVSVYCSSLACTNHNRFNPQD 122
Qy 61 SSTVOSTSETVITGTSMTGILGYDFTVQVGGISDTNQIFGLSETEPGSELYYAPFDGI 120
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
123 SSTVOSTSGTVISITGTSMTGILGYDFTVQVGGISDTNQIFGLSETEPGSELYYAPFDGI 182
Qy 121 LGLAYPSISSSGATPVFNDINWQGLVSQDLFSVYLSADDQSGSVVIFGIDSSYYTGSIN 180
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
183 LGLAYPSISSSGATPVFNDINWQGLVSQDLFSVYLSADDQSGSVVIFGIDSSYYTGSIN 242
Qy 181 WPVTVETGYQITVDSTIMNGEATACAGCAIVDTGTSLTGPTSPANTOSDITGASEN 240
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
243 WPVSVETGYQITVDSTIMNGEATACAGCAIVDTGTSLTGPTSPANTOSDITGASEN 302
Qy 241 SDGDMVVSCTAISLPDITVFTINGVQVPPPSAYILQSEGSCISGFCGMNLPTESGELWI 300
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
303 SDGDMVVSCTAISLPDITVFTINGVQVPPPSAYILQSEGSCISGFCGMNLPTESGELWI 362
Qy 301 LGDVFIRQYFTVDRANNOVGLAPVA 326
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
363 LGDVFIRQYFTVDRANNOVGLAPVA 388

RESULT 3
PEP1_MACFU
ID PEP1_MACFU STANDARD; PRT; 388 AA.
AC P03954;
DT 23-OCT-1986 (Rel. 02, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PEPsin A-1 PRECURSOR (EC 3.4.23.1) (PEPSIN III-3).
GN PGA.
OS Macaca fuscata fuscata (Japanese macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OX NCBI_TaxID=9543;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 16-70.
RC TISSUE=Gastric mucosa;
RX MEDLINE=92037645; PubMed=1935977;
RA Kageyama T., Tanabe K., Koilwai O.;
RT "Development-dependent expression of isozymogens of monkey
RT pepsinogens and structural differences between them.";
RL Eur. J. Biochem. 202:205-215(1991).
RN [2]
RP SEQUENCE OF 16-62.
RC TISSUE=Gastric mucosa;
RX MEDLINE=81006790; PubMed=6773933;
RA Kageyama T., Takahashi K.;
RT "Monkey pepsinogens and pepsins. IV. The amino acid sequence of the
RT activation peptide segment of Japanese monkey pepsinogen.";
RL J. Biochem. 88:9-16(1980).
RN [3]
RP SEQUENCE OF 41-388.
RC MEDLINE=86158132; PubMed=3514596;
RX Kageyama T., Takahashi K.;
RT "The complete amino acid sequence of monkey pepsinogen A.";
RL J. Biol. Chem. 261:4395-4405(1986).
CC -1- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
CC ALSO CLEAVED TO SOME EXTENT.
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.
CC -1- DEVELOPMENTAL STAGE: PREDOMINANT AT THE JUVENILE & ADULT STAGES.
CC -1- MISCELLANEOUS: THE EXPRESSION OF PEPsin GENES IS REGULATED BY
CC HORMONES AND RELATED SUBSTANCES.
CC -1- MISCELLANEOUS: EACH PEPsin GENES IS CONVERTED TO CORRESPONDING
CC PEPsin AT PH 2.0 IN PART AS A RESULT OF THE RELEASE OF A 47 AA
CC ACTIVATION SEGMENT AND IN PART AS A RESULT OF STEPWISE PROTEOLYTIC
CC CLEAVAGE VIA AN INTERMEDIATE FORM(S).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X59752; CAA42424.1; --
CC PIR; A00981; PEPsin.
CC PIR; S19681; S19681.
CC HSSP; P00790; IFSO.
CC MEROPS; A01.001; --
CC InterPro; IPR001461; --
CC InterPro; IPR001969; --
CC Pfam; PF00026; asp; 1.
CC PRINTS; PR00792; PEPsin.
CC PROSITE; PS00141; ASP_PROTEASE; 2.
KW Hydrolase: Aspartyl protease; Digestion; Stomach; Gastric juice;
KW Zymogen; Phosphorylation; Multigene family; Signal.
FT SIGNAL 1 15
FT PROPEP 16 40
FT CHAIN 63 388
FT DISULFID 107 112
FT DISULFID 268 272
FT DISULFID 311 344
FT ACT_SITE 94 94
FT ACT_SITE 277 277

```
FT MOD_RES 130 130 PHOSPHORYLATION.
FT CONFLICT 262 262 N -> D (IN REF. 3).
SQ SEQUENCE 388 AA; 41623 MW; 48C49BA69FD7516 CRC64;

Query Match 96.5%; Score 1648; DB 1; Length 388;
Best Local Similarity 95.1%; Pred. No. 9.7e-113;
Matches 310; Conservative 13; Mismatches 3; Indels 0; Gaps 0;

QY 1 VDEQPLENYLDMVEFGTIGTGAQDFTVVFDGSSNLWVPSVYCSLACTNHNRPED 60
DB 63 IDEQPLENYLDVEFGTIGTGAQDFTVFDGSSNLWVPSVYCSLACTNHNLPDQ 122
QY 61 SSTYQSTSETVITYGTGSMTGILGYDTVQVGGISDNTNQIFGLSETPGSLYAPDGI 120
DB 123 SSTYQSTSETVITYGTGSMTGILGYDTVQVGGISDNTNQIFGLSETPGSLYAPDGI 182
QY 121 LGLAYPSISSSGATPVFDNIWQGLVSDQLFSVYLSADDSGSGVIFGGIDSSYYTGS LN 180
DB 183 LGLAYPSISSSGATPVFDNIWQGLVSDQLFSVYLSADDSGSGVIFGGIDSSYYTGS LN 242
QY 181 WVPVTVGEGYQWITVDSITMNGEATACAEQCAIVDTGTSLLTGTPSIANIQSDIGASEN 240
DB 243 WVPVTVGEGYQWISVDSITMNGEATACAEQCAIVDTGTSLLTGTPSIANIQSDIGASEN 302
QY 241 SDGDMVYSCSAISSLPDITFTINGVQVPPPSAVILOSEGSCISGFGMNLPTESGELWI 300
DB 303 SDGDMVYSCSAISSLPDITFTINGVQVPPPSAVILOSEGSCISGFGMNDVPTESGELWI 362
QY 301 LGDVFIQRYFTVFDNRANNOVGLAPVA 326
DB 363 LGDVFIQRYFTVFDNRANNOVGLAPVA 388

RESULT 4
PEPA_MACMU STANDARD; PRT; 388 AA.
AC P11489;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PEPSIN A PRECURSOR (EC 3.4.23.1).
GN PGA.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC Cercopitheciinae; Macaca.
OX NCBI_TaxID=9544;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88313666; PubMed=2900796;
RA Evers M.P.J., Zelle B., Bebelman J.P., Pronk J.C., Mager W.H.,
RA Plantinga R.J., Eriksson A.W., Francis R.R.;
RA "Cloning and sequencing of rhesus monkey pepsinogen A cDNA.";
RL Gene 65:179-185(1988).
RN [2]
RP REVISION.
RA Zelle B.;
RL Submitted (SEP-1988) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
CC ALSO CLEAVED TO SOME EXTENT.
CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC EMBL; M20788; AAA36902.1; -.
DR PIR; J03039; PEMAOR.
DR HSSP; P00750; LPSO.
DR MEROPS; A01.001; -.
DR InterPro; IPR001461; -.
DR InterPro; IPR001969; -.
DR Pfam; PF00026; asp; 1.
DR PRINTS; PR00792; PEPsin.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
KW Hydrolase; Aspartyl protease; Digestion; Stomach; Gastric juice;
FT Zymogen; Signal; Phosphorylation.
FT SIGNAL 1 15
FT PROPEP 16 62 ACTIVATION PEPTIDE.
FT CHAIN 63 388 PEPsin A.
FT MOD_RES 130 130 PHOSPHORYLATION (POTENTIAL).
FT ACT_SITE 94 94
FT ACT_SITE 277 277
FT DISULFID 107 112 BY SIMILARITY.
FT DISULFID 268 272 BY SIMILARITY.
FT DISULFID 311 344 BY SIMILARITY.
SQ SEQUENCE 388 AA; 41696 MW; 97F6E5E3F6C2A793 CRC64;

Query Match 96.4%; Score 1646; DB 1; Length 388;
Best Local Similarity 95.1%; Pred. No. 1.4e-112;
Matches 310; Conservative 13; Mismatches 3; Indels 0; Gaps 0;

QY 1 VDEQPLENYLDMVEFGTIGTGAQDFTVVFDGSSNLWVPSVYCSLACTNHNRPED 60
DB 63 IDEQPLENYLDVEFGTIGTGAQDFTVFDGSSNLWVPSVYCSLACTNHNLPDQ 122
QY 61 SSTYQSTSETVITYGTGSMTGILGYDTVQVGGISDNTNQIFGLSETPGSLYAPDGI 120
DB 123 SSTYQSTSETVITYGTGSMTGILGYDTVQVGGISDNTNQIFGLSETPGSLYAPDGI 182
QY 121 LGLAYPSISSSGATPVFDNIWQGLVSDQLFSVYLSADDSGSGVIFGGIDSSYYTGS LN 180
DB 183 LGLAYPSISSSGATPVFDNIWQGLVSDQLFSVYLSADDSGSGVIFGGIDSSYYTGS LN 242
QY 181 WVPVTVGEGYQWITVDSITMNGEATACAEQCAIVDTGTSLLTGTPSIANIQSDIGASEN 240
DB 243 WVPVTVGEGYQWISVDSITMNGEATACAEQCAIVDTGTSLLTGTPSIANIQSDIGASEN 302
QY 241 SDGDMVYSCSAISSLPDITFTINGVQVPPPSAVILOSEGSCISGFGMNLPTESGELWI 300
DB 303 SDGDMVYSCSAISSLPDITFTINGVQVPPPSAVILOSEGSCISGFGMDVPTESGELWI 362
QY 301 LGDVFIQRYFTVFDNRANNOVGLAPVA 326
DB 363 LGDVFIQRYFTVFDNRANNOVGLAPVA 388

RESULT 5
PEPA_MACMU STANDARD; PRT; 388 AA.
AC P27678;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE PEPSIN A-4 PRECURSOR (EC 3.4.23.1) (PEPSIN I/II).
GN PGA.
OS Macaca fuscata fuscata (Japanese macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC Cercopitheciinae; Macaca.
OX NCBI_TaxID=9543;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 16-70.
RC TISSUE=Gastric mucosa;
RX MEDLINE=92037645; PubMed=1935977;
RA Kageyama T., Tanabe K., Koizumi O.;
RA "Development-dependent expression of isozymogens of monkey
```

pepsinogens and structural differences between them.";
 Eur. J. Biochem. 202:205-215(1991).
 -1- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE ALSO CLEAVED TO SOME EXTENT.
 -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-
 -1- MISCELLANEOUS: THE EXPRESSION OF PEPsin GENES IS REGULATED BY HORMONES AND RELATED SUBSTANCES.
 -1- MISCELLANEOUS: EACH PEPsin GENES IS CONVERTED TO CORRESPONDING PEPsin AT PH 2.0 IN PART AS A RESULT OF THE RELEASE OF A 47 AA ACTIVATION SEGMENT AND IN PART AS A RESULT OF STEPWISE PROTEOLYTIC CLEAVAGE VIA AN INTERMEDIATE FORM(S).
 -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE EUKARYOTIC ASPARTYL PROTEASES FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 EMBL: X59753; CAA42425.1; -;
 PIR: S16065; S16065.
 PIR: S19682; S19682.
 HSSP: P00790; IFSO.
 MEROPS: A01.001; -;
 InterPro: IPR001461; -;
 InterPro: IPR001969; -;
 Pfam: PF00026; asp; 1.
 PRINTS: PR00792; PEPsin.
 PROSITE: PS00141; ASP_PROTEASE; 2.
 KW Hydrolase; Aspartyl protease; Digestion; Stomach; Gastric juice;
 KW Zymogen; Multigene family; Signal; Glycoprotein.
 FT SIGNAL 1 15
 FT PROPEP 16 38
 FT PROPEP 39 62
 FT CHAIN 63 388
 FT ACT_SITE 94 94
 FT ACT_SITE 277 277
 FT DISULFID 107 112
 FT DISULFID 268 272
 FT DISULFID 311 344
 FT CARBOHYD 88 88
 SQ SEQUENCE 388 AA; 41955 MW; A2923AB1F7FCDEB9 CRC64;

Query Match 93.4%; Score 1595; DB 1; Length 388;
 Best Local Similarity 92.0%; Pred. No. 6.8e-109;
 Matches 300; Conservative 19; Mismatches 7; Indels 0; Gaps 0;
 QY 1 VDEQPLENYLDMEYFGTIGTIGTPAODFTVFDGSSNLWVPSVYCSLACTNHNRFNPD 60
 DB 63 IDEQPLENLDVEYFGTIGTIGTPAODFTVFDGSSNLWVPSVYCSLACTNHNRFNPD 122
 QY 61 SSTYQSTSTVITGTSMTGILGYDVTQVGGISDTNQIFGLSETEPGFLYAPFDGI 120
 DB 123 SSTYRATSKTVSITGTSMTGILGYDVTQVGGISDTNQIFGLSETEPGFLYAPFDGI 182
 QY 121 LGLAYPSISSGATPVFDNIWNGVLVSQDLFSVLSADDSQSGSVVIFGIDSYTGSUN 180
 DB 183 LGLAYPSISSGATPVFDNIWNGVLVSQDLFSVLSADDSQSGSVVIFGIDSYTGSUN 242
 QY 181 WYPTVEGFWQITVDSITNNGEIAEACQCAIVDTGTSLLGPTSPATNIQSDIGASPN 240
 DB 243 WYPTVEGFWQITVDSITNNGEIAEACQCAIVDTGTSLLGPTSPATNIQSDIGASPN 302
 QY 241 SDGDMVWSCAISLSPDIVFTINGVQYVPPSAYILQSEGSISGFGQGNLPTSEGLMI 300
 DB 303 SDGDMVWSCAISLSPDIVFTINGVQYVPPSAYILQSEGSISGFGQGNLPTSEGLMI 362
 QY 301 LGDVFIRQYFTVFDNRANNOVGLAPVA 326

DB 363 LGDVFIRQYFTVFDNRANNOVGLAPVA 388
 RESULT 6
 PEPsin_PIG STANDARD; PRT; 386 AA.
 ID PEPsin_PIG
 AC P00791;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DE 30-MAY-2000 (Rel. 39, Last annotation update)
 DE PEPsin A PRECURSOR (EC 3.4.23.1).
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88313677; PubMed=3044927;
 RA Tsukagoshi N., Ando Y., Tomita Y., Uchida R., Takemura T.,
 RA Sasaki T., Yamagata H., Uda S., Ichihara Y., Takahashi K.;
 RT "Nucleotide sequence and expression in *Escherichia coli* of cDNA of swine pepsinogen: involvement of the amino-terminal portion of the activation peptide segment in restoration of the functional protein.";
 RT FEBS Lett. 43:207-211(1974).
 RN [2]
 RP SEQUENCE OF 60-386.
 RX MEDLINE=74299591; PubMed=4604255;
 RA Moravsek L., Kostka V.;
 RT "Complete amino acid sequence of hog pepsin.";
 RT FEBS Lett. 43:207-211(1974).
 RN [3]
 RP SEQUENCE OF 16-134.
 RX MEDLINE=74031413; PubMed=4584879;
 RA Stepanov V.M., Baranova L.A., Pugacheva I.B., Belyanova L.P.,
 RA Revina L.P., Timokhina E.A.;
 RT "N-terminal sequence of swine pepsinogen and pepsin. The site of pepsinogen activation.";
 RT Biochem. Biophys. Res. Commun. 54:1164-1170(1973).
 RN [4]
 RP SEQUENCE OF 16-56.
 RX MEDLINE=69054241; PubMed=4881358;
 RA Ong E.B., Perlmann G.E.;
 RT "The amino-terminal sequence of porcine pepsinogen.";
 RT J. Biol. Chem. 243:6104-6109(1968).
 RN [5]
 RP SEQUENCE OF 58-348.
 RX MEDLINE=75211282; PubMed=1097438;
 RA Sepulveda P., Marcinszyn J.P. Jr., Liu D., Tang J.;
 RT "Primary structure of porcine pepsin. III. Amino acid sequence of a cyanogen bromide fragment, CB2A, and the complete structure of porcine pepsin.";
 RT J. Biol. Chem. 250:5082-5088(1975).
 RN [6]
 RP PARTIAL SEQUENCE OF 1-26.
 RX MEDLINE=86059312; PubMed=2415509;
 RA Ichihara Y., Sogawa K., Takahashi K.;
 RT "Isolation of human, swine, and rat prepepsinogens and calf preprochymosin, and determination of the primary structures of their NH2-terminal signal sequences.";
 RT J. Biochem. 98:483-492(1985).
 RN [7]
 RP ACTIVE SITE.
 RX MEDLINE=69283592; PubMed=4897201;
 RA Bayliss R.S., Knowles J.R., Wybrandt G.B.;
 RT "An aspartic acid residue at the active site of pepsin. The isolation and sequence of the heptapeptide.";
 RT Biochem. J. 113:377-386(1969).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY.
 RX MEDLINE=78077917; PubMed=339692;
 RA Andreeva N.S., Gustchina A.E., Fedorov A.A., Shutskever N.E.,


```
OY 121 LGLAYPSISSGATPVFDNINNOGLVSQDLFSVYLSADDSQSVVIFGIDSSYYTGSIN 180
DB 180 LGLAYPSISASGATPVFDNINNOGLVSQDLFSVYLSADDSQSVVIFGIDSSYYTGSIN 239
OY 181 WVPVVEGYWQITVDSITMNGEAIACAGCOAIVDTGTSLLTGPTSPIA-NIOSDIGASE 239
DB 240 WVPVVEGYWQITVDSITMNGEAIACAGCOAIVDTGTSLLTGPTSPIA-NIOSDIGASE 299
OY 240 NSDGMVWSCSAISSLPDIVFTINGVQVPPVPVPSAYILQSEGCISGFGQNMNLPTESGELW 299
DB 300 NSDGMVWSCSAISSLPDIVFTINGVQVPPVPVPSAYILQSEGCISGFGQNMNLPTESGELW 359
OY 300 ILGDVFIROYFTVFDNRANNOVLAPVA 326
DB 360 ILGDVFIROYFTVFDNRANNOVLAPVA 386

RESULT 7
PEP3_RABIT
ID PEP3_RABIT STANDARD; PRT; 387 AA.
AC P27822;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PEP3IN III PRECURSOR (EC 3.4.23.1) (PEPSIN A).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91009127; PubMed=2129536;
RA Kageyama T., Tanabe K., Koiwai O.;
RT "Structure and development of rabbit pepsinogens. Stage-specific
RT zymogens, nucleotide sequences of cDNAs, molecular evolution, and
RT gene expression during development.";
RL J. Biol. Chem. 265:17031-17038(1990).
CC -!- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
CC ALSO CLEAVED TO SOME EXTENT.
CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.
CC -!- DEVELOPMENTAL STAGE: PEPSINOGENS IN GROUP I, II, AND III WHERE
CC THE PREDOMINANT ZYMOGENS AT LATE POSTNATAL STAGE.
CC -!- MISCELLANEOUS: THE EXPRESSION OF PEPSINOGEN GENES IS REGULATED BY
CC HORMONES AND RELATED SUBSTANCES.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M59237; AAA85370.1; -.
DR PIR; E38302; E38302.
DR HSSP; P00791; IPSA.
DR MEROPS; A01.001; -.
DR InterPro; IPR001461; -.
DR InterPro; IPR001969; -.
DR Pfam; PF00026; asp; 1.
DR PRINTS; PR00792; PEP3IN.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
KW Hydrolase; Aspartyl protease; Digestion; Stomach; Gastric juice.
KW zymogen; Signal; Phosphorylation; Multigene family.
FT SIGNAL 1 15
FT PROPEP 16 59 ACTIVATION PEPTIDE.
FT CHAIN 60 387 PEP3IN III.
FT MOD_RES 129 129 PHOSPHORYLATION (POTENTIAL).
FT ACT_SITE 93 93 BY SIMILARITY.
FT ACT_SITE 276 276 BY SIMILARITY.
```

```
FT DISULFID 106 111 BY SIMILARITY.
FT DISULFID 267 271 BY SIMILARITY.
FT DISULFID 310 343 BY SIMILARITY.
SQ SEQUENCE 387 AA; 41969 MW; 15A59AC81F36F9EF CRC64;

Query Match 85.1%; Score 1453; DB 1; Length 387;
Best Local Similarity 82.8%; Pred. No. 1.4e-98;
Matches 270; Conservative 32; Mismatches 24; Indels 0; Gaps 0;

OY 1 VDEQPLENYLDMEYEGTIGTIPADQFTVVDFTGSSNLWVPSVYCSLACTNHNRFNPED 60
DB 62 VPTETLENYLDTEYEGTIGTIPADQFTVVDFTGSSNLWVPSVYCSRAACSVHNQFNPED 121
OY 61 SSTYQSTSETVITYGTGSMGTGILGYDVTQVQGGISDTNQIFGLSETEPGSFYLYAPFDGI 120
DB 122 SSTFQNTSESLITYGTGSMGTGFLGYDVTQVQGGIETNQIFGLSESEPGSFYLYAPFDGI 181
OY 121 LGLAYPSISSGATPVFDNINNOGLVSQDLFSVYLSADDSQSVVIFGIDSSYYTGSIN 180
DB 182 LGLAYPSISSGATPVFDNINNOGLVSQDLFSVYLSADDSQSVVIFGIDSSYYTGSIN 241
OY 181 WVPVVEGYWQITVDSITMNGEAIACAGCOAIVDTGTSLLTGPTSPIANIOSDIGASEN 240
DB 242 WVPVVEGYWQITVDSITMNGEAIACAGCOAIVDTGTSLLTGPTSPIANIOSDIGASEN 301
OY 241 SDGDMVWSCSAISSLPDIVFTINGVQVPPVPVPSAYILQSEGCISGFGQNMNLPTESGELWI 300
DB 302 SDGDMVWSCSAISSLPDIVFTINGVQVPPVPVPSAYILQSEGCISGFGQNMNLPTESGELWI 361
OY 301 LGDVFIRQYFTVFDNRANNOVLAPVA 326
DB 362 LGDVFIRQYFTVFDNRANNOVLAPVA 387

RESULT 8
PEP4_RABIT
ID PEP4_RABIT STANDARD; PRT; 387 AA.
AC P28713;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PEP3IN II-4 PRECURSOR (EC 3.4.23.1) (PEPSIN A).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91009127; PubMed=2129536;
RA Kageyama T., Tanabe K., Koiwai O.;
RT "Structure and development of rabbit pepsinogens. Stage-specific
RT zymogens, nucleotide sequences of cDNAs, molecular evolution, and
RT gene expression during development.";
RL J. Biol. Chem. 265:17031-17038(1990).
CC -!- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
CC ALSO CLEAVED TO SOME EXTENT.
CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.
CC -!- DEVELOPMENTAL STAGE: PEPSINOGENS IN GROUP I, II, AND III WHERE
CC THE PREDOMINANT ZYMOGENS AT LATE POSTNATAL STAGE.
CC -!- MISCELLANEOUS: THE EXPRESSION OF PEPSINOGEN GENES IS REGULATED BY
CC HORMONES AND RELATED SUBSTANCES.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
DR PIR; D38302; D38302.
DR HSSP; P00791; IPSA.
DR MEROPS; A01.001; -.
DR InterPro; IPR001461; -.
DR InterPro; IPR001969; -.
DR Pfam; PF00026; asp; 1.
DR PRINTS; PR00792; PEP3IN.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
```

KW	Hydrolase; Aspartyl protease; Digestion; Stomach; Gastric juice;
KW	Zymogen; Signal; Phosphorylation; Multigene family.
FT	SIGNAL 1 15
FT	PROPEP 16 59
FT	CHAIN 60 387
FT	MOD_RES 129 129
FT	ACT_SITE 93 93
FT	ACT_SITE 276 276
FT	DISULFID 106 111
FT	DISULFID 267 271
FT	DISULFID 310 343
SQ	SEQUENCE 387 AA; 42052 MW; 21ADD07782A89585 CRC64;
Query Match 80.9%; Score 1382; DB 1; Length 387;	
Best Local Similarity 79.6%; Pred. No. 2e-93;	
Matches 257; Conservative 30; Mismatches 36; Indels 0; Gaps 0;	
QY	1 VDEOPLNLDMEYFGTIGCTPAQDFTVFDGTSSNLWVPSVYCSSLACTNHNRFNPED 60
Db	62 VSTESLENLDMEYFGTIGCTPAQDFTVFDGTSSNLWVPSVYCSSLACTNHNRFNPED 121
QY	61 SSTYQSTSETVITYGTGSMTGILGYDTVQVGGISDNTQIFGLSETEPGSLYYAPFDGI 120
Db	122 SSTYQGTSETLSITYGTGSMTGILGYDTVKVGSIEDTQIFGLSKTEPFLFAPFDGI 181
QY	121 LGLAYPSISSSGATPVFDNIWNOGLVSODLFSVYLSDDEKGLVWFGGIDSSYTTGSLN 180
Db	182 LGLAYPSISSSDATPVFDNMNNEGLVSODLFSVYLSDDEKGLVWFGGIDSSYTTGSLN 241
QY	181 WVPVTEGYWQITVDSITMNGEAIACAGCOAIVDTGTSLLTGPTSPIANTQSDIGASEN 240
Db	242 WVPVTEGYWQITVDSVINGETACADSCOAIVDTGTSLLTGPTSAISNIQSYIGASKN 301
QY	241 SDGMVWVSCSAISSLPDITVFTINGVQVPPPSAVYLQSEGSCISGFCQGNLPTESELWI 300
Db	302 LLGENVISCASIDSLPDIVFTINGIQYPLPASAYILKEDDDCTSGLEGNNVDTYTGELWI 361
QY	301 LGDVFIRQYFTVFDNRANNOVGLA 323
Db	362 LGDVFIRQYFTVFDNRANNOVGLA 384
RESULT 9	
PEP2_RABIT	
ID	PEP2_RABIT
AC	P27821; STANDARD; PRT; 387 AA.
DT	01-AUG-1992 (Rel. 23, Created)
DT	01-AUG-1992 (Rel. 23, Last sequence update)
DE	30-MAY-2000 (Rel. 39, Last annotation update)
DE	PEPSIN II-2/3 PRECURSOR (EC 3.4.23.1) (PEPSIN A).
OS	Oryctolagus cuniculus (Rabbit).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
ON	NCBI_TaxID=9986;
RX	SEQUENCE FROM N.A.
RA	MEDLINE=91009127; PubMed=2129536;
RT	Kageyama T., Tanabe K., Koiwai O.;
RT	"Structure and development of rabbit pepsinogens. Stage-specific
RT	zymogens, nucleotide sequences of cDNAs, molecular evolution, and
RL	gene expression during development."
RL	J. Biol. Chem. 265:17031-17038(1990).
CC	-1- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
CC	INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
CC	ALSO CLEAVED TO SOME EXTENT.
CC	-1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-
CC	-1- DEVELOPMENTAL STAGE: PEPSINOGENS IN GROUP I, II, AND III WHERE
CC	THE PREDOMINANT ZYMOGENS AT LATE POSTNATAL STAGE.
CC	-1- MISCELLANEOUS: THE EXPRESSION OF PEPSINOGEN GENES IS REGULATED BY
CC	HORMONES AND RELATED SUBSTANCES.
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE
CC	FAMILY A1; ALSO KNOWN AS THE
CC	EUKARYOTIC ASPARTYL PROTEASES FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: M59235; AAR5369.1; -

CC F01R: C38302; C38302.

CC HSSP: P00790; IPSO.

CC MEROPS: A01.001; -

CC InterPro: IPR001461; -

CC InterPro: IPR001969; -

CC Pfam: PF00026; asp. 1.

CC PRINTS: PR00792; PEPsin.

CC PROSITE: PS00141; ASP_PROTEASE; 2.

CC Hydrolase; Aspartyl protease; digestion; Stomach; Gastric juice;

CC Zymogen; Signal; Phosphorylation; Multigene family.

FT SIGNAL 1 15

FT PROPEP 16 59

FT CHAIN 60 387

FT MOD_RES 129 129

FT ACT_SITE 93 93

FT ACT_SITE 276 276

FT DISULFID 106 111

FT DISULFID 267 271

FT DISULFID 310 343

SQ SEQUENCE 387 AA; 42100 MW; 66FC31A3DC75891 CRC64;

Query Match 80.4%; Score 1374; DB 1; Length 387;

Best Local Similarity 78.9%; Pred. No. 7.5e-93;

Matches 255; Conservative 31; Mismatches 37; Indels 0; Gaps 0;

QY 1 VDEOPLNLDMEYFGTIGCTPAQDFTVFDGTSSNLWVPSVYCSSLACTNHNRFNPED 60

Db 62 VSTESMENLDMEYFGTIGCTPAQDFTVFDGTSSNLWVPSVYCSSLACTNHNRFNPED 121

QY 61 SSTYQSTSETVITYGTGSMTGILGYDTVQVGGISDNTQIFGLSETEPGSLYYAPFDGI 120

Db 122 SSTYQGTSETLSITYGTGSMTGILGYDTVKVGSIEDTQIFGLSKTEPFLFAPFDGI 181

QY 121 LGLAYPSISSSGATPVFDNIWNOGLVSODLFSVYLSDDEKGLVWFGGIDSSYTTGSLN 180

Db 182 LGLAYPSISSSDATPVFDNMNNEGLVSODLFSVYLSDDEKGLVWFGGIDSSYTTGSLN 241

QY 181 WVPVTEGYWQITVDSITMNGEAIACAGCOAIVDTGTSLLTGPTSPIANTQSDIGASEN 240

Db 242 WVPVTEGYWQITVDSVINGETACADSCOAIVDTGTSLLTGPTSAISNIQSYIGASKN 301

QY 241 SDGMVWVSCSAISSLPDITVFTINGVQVPPPSAVYLQSEGSCISGFCQGNLPTESELWI 300

Db 302 LLGENVISCASIDSLPDIVFTINGIQYPLPASAYILKEDDDCTSGLEGNNVDTYTGELWI 361

QY 301 LGDVFIRQYFTVFDNRANNOVGLA 323

Db 362 LGDVFIRQYFTVFDNRANNOVGLA 384

RESULT 10

PEP1_RABIT

ID PEP1_RABIT

AC P28712; STANDARD; PRT; 387 AA.

DT 01-DEC-1992 (Rel. 24, Created)

DT 01-DEC-1992 (Rel. 24, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE PEPsin II-1 PRECURSOR (EC 3.4.23.1) (PEPSIN A).

OS Oryctolagus cuniculus (Rabbit).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.

ON NCBI_TaxID=9986;

11
SEQUENCE FROM N.A.
MEDLINE=91009127; PubMed=2129536;
Kageyama T., Tanabe K., Koiwai O.;
"Structure and development of rabbit pepsinogens. Stage-specific
RT zymogens, nucleotide sequences of cDNAs, molecular evolution, and
RT gene expression during development."
J. Biol. Chem. 265:17031-17038(1990).
-!- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
CC ALSO CLEAVED TO SOME EXTENT.
CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-
CC -!- DEVELOPMENTAL STAGE: PEPSINOGENS IN GROUP I, II, AND III WHERE
CC THE PREDOMINANT ZYMOKENS AT LATE POSTNATAL STAGE.
CC -!- MISCELLANEOUS: THE EXPRESSION OF PEPSINOGEN GENES IS REGULATED BY
CC HORMONES AND RELATED SUBSTANCES.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
DR PIR: B38302; B38302.
DR HSSP: P00791; IP5A.
DR MEROPS: A01.001; -.
DR InterPro: IPR001461; -.
DR InterPro: IPR001969; -.
DR Pfam: PF00026; asp. 1.
DR PRINTS: PR00792; PEPsin.
DR PROSITE: PS00141; ASP_PROTEASE; 2.
KW Hydrolase; Aspartyl protease; Digestion; Stomach; Gastric juice;
KW Zymogen; Signal; Phosphorylation; Multigene family.
FT SIGNAL 1 15
FT PROPEP 16 59 ACTIVATION PEPTIDE.
FT CHAIN 60 387
FT MOD_RES 129 129 PHOSPHORYLATION (POTENTIAL).
FT ACT_SITE 93 93 BY SIMILARITY.
FT ACT_SITE 276 276 BY SIMILARITY.
FT DISULFID 106 111 BY SIMILARITY.
FT DISULFID 267 271 BY SIMILARITY.
FT DISULFID 310 343 BY SIMILARITY.
SQ SEQUENCE 387 AA; 42070 MW; A6EC48F715541A48 CRC64;

Query Match 78.9%; Score 1348; DB 1; Length 387;
Best Local Similarity 77.3%; Pred. No. 5.8e-91;
Matches 252; Conservative 32; Mismatches 42; Indels 0; Gaps 0;

QY 1 VDEQPLENYLDMEYFGTIGTTPAQDFTVFDGSSNLWVPSVYCSLACTNHNRFNPED 60
Db 62 VSTESLENYLDAEYFGTISIGTPPOEFTVFDGSSNLWVPSVYCSLACTNHNRFNPDD 121
QY 61 STYTOSTETVITGTSMTGILGYDVTQVQGGISDTNQIFGLSETEPGSFYLYAPPDGI 120
Db 122 SSTFOATSETLITGTSMTGILGYDVTQVQGGISDTNQIFGLSETEPGITFLVAPPDGI 181
QY 121 LGLAYPSISSSGATPVFDNINWQGLVSODLFSVYLSADDSQSVVIFGIDSSYTGSLN 180
Db 182 LGLAYPSISASDTPVFDNINWQGLVSEDLFSVYLSNCKEKGSMWFGGIDSSYTGSLN 241
QY 181 WVPVTEGVQWITVDSITMNGEIAICAEQCAIVDTGTSLTGTPTSPIANIQSDIGASEN 240
Db 242 WVPVSEGVQWITVDSITMNGEIAICAEQCAIVDTGTSLTGTPTSAISKIQSYGASKN 301
QY 241 SNGDMVSCSAISSLPDIDVFTINGVQVPPPSAYILQSEGCISGIFQGNLPTESGELWI 300
Db 302 LIGENIISCAISLSDPDVFTINNVQVPLPASAYILKEDDDCLSGFDGNNLDTSYGELWI 361
QY 301 LGDVFIRQYFTVFDNRANNOVGLAPVA 326
Db 362 LGDVFIRQYFTVFDNRANNOVGLAAAA 387

RESULT 11
PEPA_CHICK
ID PEPA_CHICK PRT: 367 AA.
AC P00793;

DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PEPsin A PRECURSOR (EC 3.4.23.1).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus
OX NCBI_TaxID=9031;
RN 11
RP SEQUENCE.
RX MEDLINE=84004412; PubMed=6617663;
RA Baudys M., Kostka V.;
RT "Covalent structure of chicken pepsinogen."
RL Eur. J. Biochem. 136:89-99(1983).
CC -!- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
CC ALSO CLEAVED TO SOME EXTENT.
CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
DR PIR: A00984; PECH.
DR HSSP: P00794; 3CMS.
DR MEROPS: A01.001; -.
DR InterPro: IPR001461; -.
DR InterPro: IPR001969; -.
DR Pfam: PF00026; asp. 1.
DR PRINTS: PR00792; PEPsin.
DR PROSITE: PS00141; ASP_PROTEASE; 2.
KW Hydrolase; Aspartyl protease; Digestion; Stomach; Zymogen;
KW Glycoprotein; Gastric juice.
FT PROPEP 1 42 ACTIVATION PEPTIDE.
FT CHAIN 43 367
FT ACT_SITE 77 77
FT ACT_SITE 260 260
FT DISULFID 113 113
FT DISULFID 90 95
FT DISULFID 251 255
FT DISULFID 290 323
SQ SEQUENCE 367 AA; 40431 MW; 0C547E7FD8F5B341 CRC64;

Query Match 64.5%; Score 1102; DB 1; Length 367;
Best Local Similarity 63.5%; Pred. No. 3.9e-73;
Matches 205; Conservative 44; Mismatches 70; Indels 4; Gaps 2;

QY 4 QPLENYLDMEYFGTIGTTPAQDFTVFDGSSNLWVPSVYCSLACTNHNRFNPEDSST 63
Db 49 EPWTNMDASYGTISIGTPQDQDFVFDGSSNLWVPSVYCSLACTNHNRFNPEDSKST 108
QY 64 YQSTSETVITGTSMTGILGYDVTQVQGGISDTNQIFGLSETEPGSFYLYAPPDGI 123
Db 109 YVSTNETVITGTSMTGILGYDVTQVQGGISDTNQIFGLSETEPGSFYLYAPPDGI 168
QY 124 AYPSISSSGATPVFDNINWQGLVSODLFSVYLSADDSQSVVIFGIDSSYTGSLNWYP 183
Db 169 AEPSSISGATPVFDNINWQGLVSODLFSVYLSADDSQSVVIFGIDSSYTGSLNWYP 228
QY 184 VVVEGVQWITVDSITMNGEIAICAEQCAIVDTGTSLTGTPTSPIANIQSDIGASEN 243
Db 229 LSAETVQWITMDRVTGNGYVACFTQCAIVDTGTSLTGTPTSPIANIQSDIGASEN 286
QY 244 DMVSCSAISSLPDIDVFTINGVQVPPPSAYILQSEGCISGIFQGNLPTESGELWILD 303
Db 287 E--ISCDIDSLPDVFTINGVQVPPPSAYILQSEGCISGIFQGNLPTESGELWILD 344
QY 304 VFIROYFTVFDNRANNOVGLAPVA 326
Db 345 VFIREYVIFDNRANNOVGLSPLS 367

RESULT 12
PEPE_CHICK

```
DB 245 VSYQGYWQISMSDSIIYNKQBIACSSGQAIIDTGTSLVAGPASPDIINDIOSAVGANQNTYG 304
QY 244 DMVSCSAISSLPDVIYTINGVQYPPPSAYILQS-EGSCISFGQGNLPTSGELWILG 302
DB 305 EYSVNCSSHILAMPDVFVIGGIQYVPALAYTONGGTCMSSEFQ-----NSSADLWILG 359
QY 303 DVFETQYFTVDFRANNOVGLA 323
DB 360 DVFETVYSIFDRANNRVGLA 380

RESULT 13
PEP1_GADMO
ID PEP1_GADMO STANDARD; PRT; 324 AA.
AC P56272;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PEPSIN IIB (EC 3.4.23.-).
OS Gadus morhua (Atlantic cod).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadoidei; Gadidae;
OC Gadus.
OX NCBI_TaxID=8049;
RN [1]
SEQUENCE, AND X-RAY CRYSTALLOGRAPHY.
TISSUE-Stomach.
RA Karlisen S., Hough E., Olsen R.L.;
RT "Structure and proposed amino-acid sequence of a pepsin from Atlantic
cod (Gadus morhua).";
RL Acta Crystallogr. D 54:32-46(1998).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
EUKARYOTIC ASPARTYL PROTEASES FAMILY.
DR PDB; 1AMS; 24-DEC-97.
DR InterPro; IPR001461; -.
DR InterPro; IPR001969; -.
DR Pfam; PF00026; asp; 1.
DR PRINTS; PR00792; PEPSIN.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
KW Hydrolase; Aspartyl protease; Digestion; Stomach; Gastric juice;
KW 3D-structure.
FT ACT_SITE 32 32 BY SIMILARITY.
FT ACT_SITE 214 214 BY SIMILARITY.
FT DISULFID 45 50 BY SIMILARITY.
FT DISULFID 206 209 BY SIMILARITY.
FT DISULFID 247 280 BY SIMILARITY.
SQ SEQUENCE 324 AA; 34014 MW; EE3A6097B6941DD7 CRC64;

Query Match 61.7%; Score 1053.5; DB 1; Length 324;
Best Local Similarity 59.8%; Pred. No. 1.1e-69;
Matches 195; Conservative 49; Mismatches 79; Indels 3; Gaps 3;

*QY 1 VDEQPLENYLDMYEGTIGTTPAODFTVFTGSSNLWVPVSYVSSSLACTNHNRFNPD 60
DB 2 VTEQ-MKNEADPEYGVISIGTTPPEFVIFDTGSSNLWSSSHCSAQACSNHNKFRQ 60
QY 61 SSTYQSTFTVITYTGTGTCMIGLYDVQVGGISDTNIFGLSETPEGSFLYAPFDGI 120
DB 61 SSTYVETGKVDLTGTGCMRGILQDQIVSVGGSDPNQELGESQTEPGFOAAAPFDGI 120
QY 121 LGLAYPSISSSGATPVFDNINWGLVSDQLFSVYLSADQSGSVIFGIDSSYTGSLN 180
DB 121 LGLAYPSIAAAGAVPFDNMGSQSLVEXDLFSYLSGGGANGSEVNLGVDNHSYTGSIH 180
QY 181 WYPTVEGWQITVDSITMNGEIAICAECCOAIQVDTGTSLLTGPTSPIANIQSDIGASEN 240
DB 181 WTPVTAEKYQWALDGTIVNGTAAEC-EGCAIVDTGTSKIVAPVAPVAPVAPVAPVAPV 239
QY 241 SDGDMVWSCAISLSLPDVIYTINGVQYPPPSAYILQS-EGSCISFGQGNLPTSGELWI 300
DB 245 VSYQGYWQISMSDSIIYNKQBIACSSGQAIIDTGTSLVAGPASPDIINDIOSAVGANQNTYG 304
```

```
DB 245 VSYQGYWQISMSDSIIYNKQBIACSSGQAIIDTGTSLVAGPASPDIINDIOSAVGANQNTYG 304
QY 244 DMVSCSAISSLPDVIYTINGVQYPPPSAYILQS-EGSCISFGQGNLPTSGELWILG 302
DB 305 EYSVNCSSHILAMPDVFVIGGIQYVPALAYTONGGTCMSSEFQ-----NSSADLWILG 359
QY 303 DVFETQYFTVDFRANNOVGLA 323
DB 360 DVFETVYSIFDRANNRVGLA 380

RESULT 13
PEP1_GADMO
ID PEP1_GADMO STANDARD; PRT; 324 AA.
AC P56272;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PEPSIN IIB (EC 3.4.23.-).
OS Gadus morhua (Atlantic cod).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadoidei; Gadidae;
OC Gadus.
OX NCBI_TaxID=8049;
RN [1]
SEQUENCE, AND X-RAY CRYSTALLOGRAPHY.
TISSUE-Stomach.
RA Karlisen S., Hough E., Olsen R.L.;
RT "Structure and proposed amino-acid sequence of a pepsin from Atlantic
cod (Gadus morhua).";
RL Acta Crystallogr. D 54:32-46(1998).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
EUKARYOTIC ASPARTYL PROTEASES FAMILY.
DR PDB; 1AMS; 24-DEC-97.
DR InterPro; IPR001461; -.
DR InterPro; IPR001969; -.
DR Pfam; PF00026; asp; 1.
DR PRINTS; PR00792; PEPSIN.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
KW Hydrolase; Aspartyl protease; Digestion; Signal; Glycoprotein.
KW SIGNAL
FT CHAIN 1 16
FT ACT_SITE 17 383 BY SIMILARITY.
FT ACT_SITE 94 94 BY SIMILARITY.
FT ACT_SITE 276 276 BY SIMILARITY.
FT DISULFID 107 112 BY SIMILARITY.
FT DISULFID 267 271 BY SIMILARITY.
FT DISULFID 310 344 BY SIMILARITY.
FT CARBOHYD 132 132 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 204 204 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 309 309 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 350 350 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 51 51 T -> S.
SQ SEQUENCE 383 AA; 41719 MW; 1642796871611F54 CRC64;

Query Match 62.6%; Score 1068.5; DB 1; Length 383;
Best Local Similarity 61.4%; Pred. No. 1.1e-70;
Matches 197; Conservative 53; Mismatches 64; Indels 7; Gaps 3;

*QY 4 QPLENYLDMYEGTIGTTPAODFTVFTGSSNLWVPVSYVSSSLACTNHNRFNPDSS 63
DB 66 EPLLTLDMEYGTISIGTTPPDFTVFTGSSNLWVPVSVCTSPACQSHQMFNPSSQST 125
QY 64 YQSTSETVITYTGTGTCMIGLYDVQVGGISDTNIFGLSETPEGSFLYAPFDGI 123
DB 126 YKSTGQNLSTHYGTGDMGTGVCDDTVTVASLMDTNLFLGLSTSEPGQFFVYVKFDGILGL 185
QY 124 AYPSSISGATPVFDNINWGLVSDQLFSVYLSADQSGSVIFGIDSSYTGSLNWPV 183
DB 186 GYPSLAADGTPVFDNWNESLEQNLFVYLSREPW-GSMVFGGIDSEYFTGSLNWP 244
QY 184 VTVEGYWQITVDSITMNGEIAICAECCOAIQVDTGTSLLTGPTSPIANIQSDIGASENSDG 243
```

Db 240 -GEMMNCASVQSPLDPTFTTNGVQKPLPPSAYIEGQAFCTSGLGSSGVPSNTSELWI 298

Qy 301 LGDFVFIROYFTVDRANNOVGLAPVA 326

Db 299 FGDVFLRNYTYIYDRTNKNKVGAPAA 324

RESULT 14

ID PEPP_RABIT STANDARD; PRT; 388 AA.
AC P27823;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PEPSIN F PRECURSOR (EC 3.4.23.1).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91009127; PubMed=2129536;
RA Kageyama T., Tanabe K., Koiwai O.;
RT "Structure and development of rabbit pepsinogens. Stage-specific
RT zymogens, nucleotide sequences of cDNAs, molecular evolution, and
RT gene expression during development.";
RL J. Biol. Chem. 265:17031-17038(1990).
CC -!- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
CC ALSO CLEAVED TO SOME EXTENT.
CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.
CC -!- DEVELOPMENTAL STAGE: EARLY POSTNATAL.
CC -!- MISCELLANEOUS: THE EXPRESSION OF PEPSINOGEN GENES IS REGULATED BY
CC HORMONES AND RELATED SUBSTANCES.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M59238; AAA31440.1; -;
DR PIR; A38302; A38302.
DR HSP; P00794; 3CMS.
DR InterPro; IPR001461; -;
DR InterPro; IPR001969; -;
DR Pfam; PF00026; asp; 1.
DR PRINTS; PR00792; PEPSIN.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
KW Hydrolase; Aspartyl protease; Digestion; Stomach; Gastric juice;
KW Zymogen; Signal; Multigene family.
FT SIGNAL 1 15
FT PROPEP 16 58
FT CHAIN 59 388
FT ACT_SITE 92 92
FT ACT_SITE 275 275
FT DISULFID 105 110
FT DISULFID 266 270
FT DISULFID 309 343
FT SEQUENCE 388 AA; 42786 MW; 24792BE393594B3A CRC64;

Query Match 60.3%; Score 1030; DB 1; Length 388;
Best Local Similarity 58.3%; Pred. No. 7.le-68;
Matches 190; Conservative 49; Mismatches 85; Indels 2; Gaps 2;

Qy 1 VDEQPLEYLDMEYFGTIGTGPADFTVVDYTGSSNLWVPSVYGCSSLACTIONHNFNPD 60

Db 61 VDFEPLRNYLDIAYIGTIGTPPEEFKVLDTGSADLWVPSIYCSSPACGKHNTFNPLL 120

Qy 61 SSTYSTSETSYTGTGSMGTGILGYDTVQVGGISDTNQIFGLSETEPSEFLYYAPFDGI 120
Db 121 SSTFLVSGRPINIVYSGRMSGLAYDTVQIAGLVDAQAFGLSLQEPKPMFYAVFDGI 180
Qy 121 LGLAYPSISSSGATPVFDNIWNOGLVSODLFSVLSADDDQSGSVVIFGIDSSYTTGSLN 180
Db 181 LGLSYPSLSFEGITPVFDNLMAQGLISQNLFAFYLSKKEERGSMMLGVDVPSYYSGDLH 240
Qy 181 WVPVTVEGYWQITVDSITMNGEATACAGCOAIVDTGTSLTGTPTSPIANQSDIGASEN 240
Db 241 WVPVSRPLYWQLAVDRISMNGEATCGDCGCGIVDTGTSLIGPRDPVLNOKIINAQHS 300
Qy 241 SDGDMVSCSAISSLPDIVFTTINGVQYVPPSPAYTLQSE-GSCISGF-QGMNLPTESEL 298
Db 301 HGGEVIIDCDTISTLPDIIFTIDGVDPVPASAYIRKSSVHGVCYNFDESAHAHESEPEV 360
Qy 299 WILGDVFIROYFTVDRANNOVGLAP 324
Db 361 WVLGDVFLRLYFTVDRANNRIGLAP 386

RESULT 15

ID PAG_HORSE STANDARD; PRT; 388 AA.
AC Q28389;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PREGNANCY-ASSOCIATED GLYCOPROTEIN PRECURSOR (EC 3.4.23.-) (PAG).
GN PAG.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Placenta;
RX MEDLINE=98221985; PubMed=9561214;
RA Green J., Xie S., Gan X., Roberts R.M.;
RT "An aspartic proteinase expressed in the equine placenta.";
RL Adv. Exp. Med. Biol. 436:163-167(1998).
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -!- TISSUE SPECIFICITY: TROPHOBLAST AND PLACENTAL TISSUE.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; L38511; AAC14885.1; -;
DR HSP; P20142; IAVF.
DR MEROPS; A01.091; -;
DR InterPro; IPR001461; -;
DR InterPro; IPR001969; -;
DR Pfam; PF00026; asp; 1.
DR PRINTS; PR00792; PEPSIN.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
KW Hydrolase; Aspartyl protease; Glycoprotein; Signal; Multigene family;
KW Zymogen.

FT SIGNAL 1 15
FT PROPEP 16 92
FT CHAIN 92 388
FT ACT_SITE 92 92
FT ACT_SITE 275 275
FT DISULFID 105 110
FT DISULFID 266 270
FT DISULFID 309 344

POTENTIAL.
ACTIVATION PEPTIDE (POTENTIAL).
PREGNANCY-ASSOCIATED GLYCOPROTEIN.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OW protein - protein search, using sw model

Run on: September 6, 2001, 16:45:56 ; Search time 134.15 Seconds
(without alignments)
185.113 Million cell updates/sec

Title: US-09-603-713-31
Perfect score: 1708
Sequence: 1 VDEQPLENLDMEYFETIGI.....ROYFTVFDNRANNQVGLAPVA 326

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1708	100.0	388	1 PEHU	pepsin A (EC 3.4.2
2	1701	99.6	388	2 A30142	pepsin A (EC 3.4.2
3	1690	98.9	388	2 B30142	pepsin A (EC 3.4.2
4	1672	97.9	388	1 S19684	pepsin A (EC 3.4.2
5	1648	96.5	388	1 PEMQAJ	pepsin A (EC 3.4.2
6	1646	96.4	388	1 PEMQAR	pepsin A (EC 3.4.2
7	1595	93.4	388	1 S19682	pepsin A (EC 3.4.2
8	1558	91.2	387	2 JC7245	pepsinogen A - com
9	1508.5	88.3	386	1 PEPG	pepsin A (EC 3.4.2
10	1453	85.1	387	2 E38302	pepsin (EC 3.4.23.
11	1382	80.9	387	2 D38302	pepsin (EC 3.4.23.
12	1374	80.4	387	2 D38302	pepsin (EC 3.4.23.
13	1348	78.9	387	2 B38302	pepsin (EC 3.4.23.
14	1311	76.8	384	3 JC7574	pepsinogen A - Afr
15	1304	76.3	385	3 JC7575	pepsinogen A - bul
16	1249	73.1	334	2 JC4870	pepsin A (EC 3.4.2
17	1106	64.8	382	1 PECH	pepsin A (EC 3.4.2
18	1068.5	62.6	383	2 A41443	prochymosin - comm
19	1041.5	61.0	381	2 JC7247	chymosin (EC 3.4.23.
20	1021.5	59.8	380	2 A47176	chymosin (EC 3.4.23.
21	1019.5	59.7	389	2 A38302	chymosin (EC 3.4.23.
22	1018.5	59.6	381	1 CMSHB	chymosin (EC 3.4.2
23	1008.5	59.0	381	1 CMBO	chymosin (EC 3.4.2
24	988.5	57.9	396	2 A34401	cathepsin E (EC 3.4
25	985.5	57.7	384	2 A39314	gastricsin (EC 3.4
26	980	57.4	383	3 JC7573	pepsinogen C - Afr
27	972.5	56.9	398	2 S66465	cathepsin E (EC 3.
28	967.5	56.1	396	2 S36855	cathepsin E (EC 3.
29	930	54.4	391	2 A43356	cathepsin E (EC 3.

ALIGNMENTS

RESULT 1

PEHU

pepsin A (EC 3.4.23.1) 3 precursor [validated] - human

N:Alternate names: pepsinogen A isozyme 3

C:Species: Homo sapiens (man)

C:Date: 19-Feb-1984 #sequence_revision 19-Feb-1984 #text_change 08-Dec-2000

C:Accession: A00980; PX0023; S02663; F2434; I54252; PX0024

R:Sogawa, K.; Fujii-Kuriyama, Y.; Mizukami, Y.; Ichihara, Y.; Takahashi, K.

J. Biol. Chem. 258, 5306-5311, 1983

A:Title: Primary structure of human pepsinogen gene.

A:Reference number: A00980; MUID:83161158

A:Accession: A00980

A:Molecule type: DNA

A:Residues: 1-388 <SOG>

A:Cross-references: GB:J00279

R:Atanada, S.B.P.; Tanji, M.; Kageyama, T.; Takahashi, K.

J. Biochem. 106, 920-927, 1989

A:Title: A comparative study on the NH2-terminal amino acid sequences and some other

A:Reference number: PX0023; MUID:90130402

A:Accession: PX0023

A:Molecule type: protein

A:Residues: 16-100 <ATH>

R:Foltmann, B.

FEBS Lett. 241, 69-72, 1988

A:Title: Activation of human pepsinogens.

A:Reference number: S02663; MUID:89065108

A:Accession: S02663

A:Molecule type: protein

A:Residues: 16-68 <FOL>

R:Ichihara, Y.; Sogawa, K.; Takahashi, K.

J. Biochem. 98, 483-492, 1985

A:Title: Isolation of human, swine, and rat prepepsinogens and calf preprochymosin, a

A:Reference number: A22434; MUID:86059312

A:Accession: F22434

A:Molecule type: protein

A:Residues: 1-15, 'XX', 19-20, 'X', 22, 'XX', 25-26, 'X', 28 <ICH>

R:Evers, M.P.J.; Zelle, B.; Peepker, D.S.; Mager, W.H.; Planta, R.J.; Eriksson, A.W.;

Hum. Genet. 77, 182-187, 1987

A:Title: Molecular cloning of a pair of human pepsinogen A genes which differ by a G1

A:Reference number: I54252; MUID:88006181

A:Accession: I54252

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-27, 'F', 29-73 <RES>

A:Cross-references: GB:M27598; NID:g189834; PIDN:AAA36431.1; PID:g189836

C:Genetics:

A:Gene: GDB:PGAS3

A:Cross-references: GDB:I19482; OMIM:169710

A:Map position: 11q13.1-11q13.5

A:Introns: 19/2; 73/3; 113/1; 152/3; 219/2; 258/2; 306/3; 339/3

C:Superfamily: pepsin

C:Keywords: aspartic proteinase; gastric juice; hydrolase; phosphoprotein; protein di

F;1-15/Domain: signal sequence #status experimental <SIG>
F;16-388/Product: pepsinogen A 3 #status experimental <ZYM>
F;16-62/Domain: activation peptide #status experimental <APT>
F;60-388/Product: pepsin A 3, minor variant #status experimental <MIN>
F;63-388/Product: pepsin A 3 #status experimental <MAT>
F;94,277/Active site: Asp #status predicted
F;130/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 100.0%; Score 1708; DB 1; Length 388;
Best Local Similarity 100.0%; Pred. No. 7.3e-115;
Matches 326; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VDEQPLENYLDMYFGTIGTGAQDFTVVFDGSSNLWVPSVYCSSLACTNHNRFNPED 60
Db 63 VDEQPLENYLDMYFGTIGTGAQDFTVVFDGSSNLWVPSVYCSSLACTNHNRFNPED 122

Qy 61 SSTVQSTSETVITYGTSGMTGILGYDTVQVGGISDNTQIFGLSETPGSLYAPFDGI 120
Db 123 SSTVQSTSETVITYGTSGMTGILGYDTVQVGGISDNTQIFGLSETPGSLYAPFDGI 182

Qy 121 LGLAYPSISSSGATPVFDNTWNGQLVSQDLFSVYLSADDSQGSVVFIFGGIDSSYYTGS LN 180
Db 183 LGLAYPSISSSGATPVFDNTWNGQLVSQDLFSVYLSADDSQGSVVFIFGGIDSSYYTGS LN 242

Qy 181 WVPVTVEGYWQITVDSITMNGEAIACAEQCOAIVDTGTSLTGTPTSPANTQSDIGASEN 240
Db 243 WVPVTVEGYWQITVDSITMNGEAIACAEQCOAIVDTGTSLTGTPTSPANTQSDIGASEN 302

Qy 241 SDGDMVYSCSAISSLPDITFTINGVQVPPPSAVILOSEGSISGFQGMNLPTESELWI 300
Db 303 SDGDMVYSCSAISSLPDITFTINGVQVPPPSAVILOSEGSISGFQGMNLPTESELWI 362

Qy 301 LGDVFIRQYFTVFDNRANNOVGLAPVA 326
Db 363 LGDVFIRQYFTVFDNRANNOVGLAPVA 388

RESULT 2

A30142
Pepsin A (EC 3.4.23.1) 5 precursor - human
N;Alternate names: pepsinogen 5
C;Species: Homo sapiens (man)
C;Date: 07-Jun-1990 #sequence_revision 20-Aug-1994 #text_change 29-Aug-1997
C;Accession: A30142; S02664; S02542; PX0027; PX0025; PX0026; A22434
R;Evers, M.P.J.; Zelle, B.; Bebelman, J.P.; van Beusechem, V.; Kraakman, L.; Hoffer, M.J.
Genomics 4, 232-239, 1989
A;Title: Nucleotide sequence comparison of five human pepsinogen A (PGA) genes: evolution
A;Reference number: A91627; MUID:89233110
A;Accession: A30142
A;Molecule type: DNA
A;Residues: 1-27, 'F', 29-388 <EVE>
A;Cross-references: GB:M26025
A;Note: the authors translated the codon TTC for residue 28 as Leu, GGC for residue 36 as
R;Foltmann, B.
FEBS Lett. 241, 69-72, 1988
A;Title: Activation of human pepsinogens.
A;Reference number: S02663; MUID:89065108
A;Accession: S02664
A;Molecule type: protein
A;Residues: 16-68 <FOL>
R;Bank, R.A.; Crusius, B.C.; Zwiers, T.; Meunissen, S.G.M.; Arwert, F.; Pronk, J.C.
FEBS Lett. 238, 105-108, 1988
A;Title: Identification of a Glu > Lys substitution in the activation segment of human p
A;Reference number: S02542; MUID:89005649
A;Accession: S02542
A;Status: preliminary
A;Molecule type: protein
A;Residues: 16-58 <BAN>
R;Athauda, S.B.P.; Tanji, M.; Kageyama, T.; Takahashi, K.
J. Biochem. 106, 920-927, 1989
A;Title: A comparative study on the NH2-terminal amino acid sequences and some other pro
A;Reference number: PX0023; MUID:90130402

A;Accession: PX0027
A;Molecule type: protein
A;Residues: 16-100 <ATH>
A;Accession: PX0025
A;Molecule type: protein
A;Residues: 16-60 <AT2>
A;Accession: PX0026
A;Molecule type: protein
A;Residues: 16-100 <AT3>
R;Ichihara, Y.; Sogawa, K.; Takahashi, K.
J. Biochem. 98, 483-492, 1985
A;Title: Isolation of human, swine, and rat prepepsinogens and calf preprochymosin, a
A;Reference number: A22434; MUID:86059312
A;Accession: A22434
A;Molecule type: protein
A;Residues: 1-15, 'XXX', 19-20, 'X', 22, 'XX', 25-26, 'X', 28 <ICH>
C;Genetics:

A;Gene: GDB:PGA5
A;Cross-references: GDB:119484; OMIM:169730
A;Map position: 11q13-11q13
A;Introns: 19/1; 73/3; 113/1; 152/3; 219/2; 258/2; 306/3; 339/3
C;Superfamily: pepsin

C;Keywords: aspartic proteinase; hydrolase; phosphoprotein; protein digestion
F;1-15/Domain: signal sequence #status experimental <SIG>
F;16-388/Product: pepsinogen A 5 #status experimental <ZYM>
F;60-388/Product: pepsin A 5, minor variant #status experimental <MIN>
F;63-388/Product: pepsin A 5 #status experimental <MAT>
F;94,277/Active site: Asp #status predicted
F;130/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 99.6%; Score 1701; DB 2; Length 388;
Best Local Similarity 99.4%; Pred. No. 2.3e-114;
Matches 324; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VDEQPLENYLDMYFGTIGTGAQDFTVVFDGSSNLWVPSVYCSSLACTNHNRFNPED 60
Db 63 VDEQPLENYLDMYFGTIGTGAQDFTVVFDGSSNLWVPSVYCSSLACTNHNRFNPED 122

Qy 61 SSTVQSTSETVITYGTSGMTGILGYDTVQVGGISDNTQIFGLSETPGSLYAPFDGI 120
Db 123 SSTVQSTSETVITYGTSGMTGILGYDTVQVGGISDNTQIFGLSETPGSLYAPFDGI 182

Qy 121 LGLAYPSISSSGATPVFDNTWNGQLVSQDLFSVYLSADDSQGSVVFIFGGIDSSYYTGS LN 180
Db 183 LGLAYPSISSSGATPVFDNTWNGQLVSQDLFSVYLSADDSQGSVVFIFGGIDSSYYTGS LN 242

Qy 181 WVPVTVEGYWQITVDSITMNGEAIACAEQCOAIVDTGTSLTGTPTSPANTQSDIGASEN 240
Db 243 WVPVTVEGYWQITVDSITMNGEAIACAEQCOAIVDTGTSLTGTPTSPANTQSDIGASEN 302

Qy 241 SDGDMVYSCSAISSLPDITFTINGVQVPPPSAVILOSEGSISGFQGMNLPTESELWI 300
Db 303 SDGDMVYSCSAISSLPDITFTINGVQVPPPSAVILOSEGSISGFQGMNLPTESELWI 362

Qy 301 LGDVFIRQYFTVFDNRANNOVGLAPVA 326
Db 363 LGDVFIRQYFTVFDNRANNOVGLAPVA 388

RESULT 3

B30142
Pepsin A (EC 3.4.23.1) 4 precursor - human
C;Species: Homo sapiens (man)
C;Date: 07-Jun-1990 #sequence_revision 20-Aug-1994 #text_change 29-Aug-1997
C;Accession: B30142; E22434
R;Evers, M.P.J.; Zelle, B.; Bebelman, J.P.; van Beusechem, V.; Kraakman, L.; Hoffer,
Genomics 4, 232-239, 1989
A;Title: Nucleotide sequence comparison of five human pepsinogen A (PGA) genes: evolu
A;Reference number: A91627; MUID:89233110
A;Accession: B30142
A;Molecule type: DNA
A;Residues: 1-27, 'F', 29-388 <EVE>

A:Note: the authors translated the codon TTC for residue 28 as Leu, GGC for residue 36 as
 R:Ichihara, Y.; Sogawa, K.; Takahashi, K.

J. Biochem. 98, 483-492, 1995

A:Title: Isolation of human, swine, and rat prepepsinogens and calf preprochymosin, and
 A:Reference number: A22434; MUID:86059312

A:Accession: E22434

A:Molecule type: protein

A:Residues: 1-15, 'XX', 19-20, 'X', 22, 'XX', 25-26, 'X', 28 <ICH>

C:Genetics:

A:Gene: GDB:PG44

A:Cross-references: GDB:119483; OMIM:169720

A:Map position: 11q13-11q13

A:Introns: 19/1; 73/3; 113/1; 152/3; 219/2; 258/2; 306/3; 339/3

C:Superfamily: pepsin

C:Keywords: aspartic proteinase; hydrolase; phosphoprotein; protein digestion; zymogen

F:1-15/Domain: signal sequence #status experimental <SIG>

F:16-59/Domain: activation peptide #status experimental <APT>

F:63-388/Product: pepsin A 4 #status predicted <MAT>

F:94,277/Active site: Asp #status predicted

F:107-112,268-272,311-344/Disulfide bonds: #status predicted

F:130/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 98.9%; Score 1690; DB 2; Length 388;
 Best Local Similarity 98.5%; Pred. No. 1.4e-113;
 Matches 321; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VDEQPLENLDMEYFGTIGTGAQDFVFDTSNLLWVPSVYCSLACTNHNRFNPD 60

Db 63 VDEQPLENLDMEYFGTIGTGAQDFVFDTSNLLWVPSVYCSLACTNHNRFNPD 122

Qy 61 SSTQSTSETVITYGTGSMTGILGYDVTQVGGISDTNQIFGLSETEPGSFLLYAPFDGI 120

Db 123 SSTQSTSETVITYGTGSMTGILGYDVTQVGGISDTNQIFGLSETEPGSFLLYAPFDGI 182

Qy 121 LGLAYPSISSGATPVFDNIWQGLVSQDLFSVLSADDSGVSVIFGIDSSYYTGSIN 180

Db 183 LGLAYPSISSGATPVFDNIWQGLVSQDLFSVLSADDSGVSVIFGIDSSYYTGSIN 242

Qy 181 WPVTVGEGWQITVDSITMNGEATACAGCOAIVDTGTSLLTGTPTSPANTOSDIGASEN 240

Db 243 WPVTVGEGWQITVDSITMNGEATACAGCOAIVDTGTSLLTGTPTSPANTOSDIGASEN 302

Qy 241 SDGMVWSCSAISSLPDITVFTINGVQYVPVPSAYILQSEGSCISGFGQNMNLPTESELGI 300

Db 303 SDGMVWSCSAISSLPDITVFTINGVQYVPVPSAYILQSEGSCISGFGQNMNLPTESELGI 362

Qy 301 LGDVFIRQYFTVFRANNOVGLAPVA 326

Db 363 LGDVFIRQYFTVFRANNOVGLAPVA 388

RESULT 4
 SI9684
 Pepsin A (EC 3.4.23.1) 2/3 precursor - Japanese macaque
 A:Alternate names: pepsinogen A isozyme 2/3
 C:Species: Macaca fuscata (Japanese macaque)
 C:Date: 22-Nov-1993 #sequence_revision 19-Oct-1995 #text_change 18-Jun-1999
 C:Accession: SI9684; SI6064
 R:Kageyama, T.; Tanabe, K.; Koikawa, O.
 Eur. J. Biochem. 202, 205-215, 1991

A:Title: Development-dependent expression of isozymogens of monkey pepsinogens and struc
 A:Reference number: SI9681; MUID:92037645
 A:Accession: SI9684
 A:Molecule type: mRNA

A:Residues: 1-388 <KAG>

A:Cross-references: EMBL:X59755; NID:g38068; PIDN:CAA42427.1; PID:g38069

A:Note: parts of sequence, including amino ends of pepsinogen and activation intermediat
 C:Comment: It could not be determined if this sequence represents isozyme 2 or 3, which
 in by 4 months of age.

C:Comment: Although two-step activation is observed, activation is predominantly a one-
 C:Superfamily: pepsin

C:Keywords: aspartic proteinase; gastric juice; hydrolase; phosphoprotein; protein diges

F:1-15/Domain: signal sequence #status predicted <SIG>

F:16-388/Product: pepsinogen A 2/3 #status experimental <APT>

F:63-388/Product: pepsin A 2/3 #status experimental <APT>

F:40-41/Cleavage site: Asp-Phe (pepsin) #status experimental

F:62-63/Cleavage site: Leu-Ile (pepsin) #status experimental

F:94,277/Active site: Asp #status predicted

F:107-112,268-272,311-344/Disulfide bonds: #status predicted

F:130/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 97.9%; Score 1672; DB 1; Length 388;
 Best Local Similarity 96.6%; Pred. No. 2.7e-112;
 Matches 315; Conservative 10; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VDEQPLENLDMEYFGTIGTGAQDFVFDTSNLLWVPSVYCSLACTNHNRFNPD 60

Db 63 IDEQPLENLDMEYFGTIGTGAQDFVFDTSNLLWVPSVYCSLACTNHNRFNPD 122

Qy 61 SSTQSTSETVITYGTGSMTGILGYDVTQVGGISDTNQIFGLSETEPGSFLLYAPFDGI 120

Db 123 SSTQSTSETVITYGTGSMTGILGYDVTQVGGISDTNQIFGLSETEPGSFLLYAPFDGI 182

Qy 121 LGLAYPSISSGATPVFDNIWQGLVSQDLFSVLSADDSGVSVIFGIDSSYYTGSIN 180

Db 183 LGLAYPSISSGATPVFDNIWQGLVSQDLFSVLSADDSGVSVIFGIDSSYYTGSIN 242

Qy 181 WPVTVGEGWQITVDSITMNGEATACAGCOAIVDTGTSLLTGTPTSPANTOSDIGASEN 240

Db 243 WPVTVGEGWQITVDSITMNGEATACAGCOAIVDTGTSLLTGTPTSPANTOSDIGASEN 302

Qy 241 SDGMVWSCSAISSLPDITVFTINGVQYVPVPSAYILQSEGSCISGFGQNMNLPTESELGI 300

Db 303 SDGMVWSCSAISSLPDITVFTINGVQYVPVPSAYILQSEGSCISGFGQNMNLPTESELGI 362

Qy 301 LGDVFIRQYFTVFRANNOVGLAPVA 326

Db 363 LGDVFIRQYFTVFRANNOVGLAPVA 388

RESULT 5
 PEMQAJ
 Pepsin A (EC 3.4.23.1) 1 precursor - Japanese macaque
 A:Alternate names: pepsinogen A isozyme 1
 C:Species: Macaca fuscata (Japanese macaque)
 C:Date: 13-Aug-1986 #sequence_revision 19-Oct-1995 #text_change 18-Jun-1999
 C:Accession: SI9581; A92579; A00981
 R:Kageyama, T.; Tanabe, K.; Koikawa, O.
 Eur. J. Biochem. 202, 205-215, 1991
 A:Title: Development-dependent expression of isozymogens of monkey pepsinogens and st
 A:Reference number: SI9681; MUID:92037645
 A:Accession: SI9681
 A:Molecule type: mRNA

A:Residues: 1-388 <KAG>

A:Cross-references: EMBL:X59752; NID:g38074; PIDN:CAA42424.1; PID:g38075

A:Note: parts of sequence, including amino ends of pepsinogen and activation intermed

R:Kageyama, T.; Takahashi, K.

J. Biochem. 88, 9-16, 1980

A:Title: Monkey pepsinogens and pepsins. IV. The amino acid sequence of the activatio

A:Reference number: A91960; MUID:81006790

A:Accession: A91960

A:Molecule type: protein

A:Residues: 16-62 <KAZ>

R:Kageyama, T.; Takahashi, K.

J. Biol. Chem. 261, 4395-4405, 1986

A:Title: The complete amino acid sequence of monkey pepsinogen A.

A:Reference number: A92579; MUID:86168132

A:Accession: A92579

A:Molecule type: protein

A:Residues: 41-261, 'D', 263-388 <KAZ>

C:Comment: This is the major pepsin isozyme in juveniles and adults.

C:Comment: Activation is a one-step process.

C:Superfamily: pepsin

C:Keywords: aspartic proteinase; gastric juice; hydrolase; phosphoprotein; protein digest
 F:1-15/Domain: signal sequence #status predicted <SIG>
 F:16-62/Domain: activation peptide #status experimental <APT>
 F:63-388/Product: pepsin A 1 #status experimental <ENZ>
 F:94,277/Active site: Asp #status predicted
 F:107-112,268-272,311-344/Disulfide bonds: #status experimental
 F:130/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match 96.5%; Score 1648; DB 1; Length 388;
 Best Local Similarity 95.1%; Pred. No. 1.4e-110;
 Matches 310; Conservative 13; Mismatches 3; Indels 0; Gaps 0;

Qy 1 VDEQPLENLDMEYFGTIGTIGTTPAQDFTVFDTSNNLWVPSVYCSLACTNHNRFNPD 60
 Db 63 IDEQPLENLDVEYFGTIGTIGTTPAQDFTVFDTSNNLWVPSVYCSLACTNHNLFNPD 122
 Qy 61 SSTQSTSETSVITYGTGSMTGILGYDTPVQGGISDTNQIFGLSETEPGSFLYYAPFDGI 120
 Db 123 SSTQSTSETSVITYGTGSMTGILGYDTPVQGGISDTNQIFGLSETEPGSFLYYAPFDGI 182
 Qy 121 LGLAYPSISSSGATPVFDNIWQGLVSODLFSVYLSADDDQSGSVVIFGGIDSSYYTGS LN 180
 Db 183 LGLAYPSISSSGATPVFDNIWQGLVSODLFSVYLSADDDQSGSVVIFGGIDSSYYTGS LN 242
 Qy 181 WVPVTVEGYWQITVDSITMNGEATACAGCOAIVDTGTSLLTGTPSPANTQSDIGASEN 240
 Db 243 WVPVSVEGYWQISVDSITMNGEATACAGCOAIVDTGTSLLTGTPSPANTQSDIGASEN 302
 Qy 241 SDGDMVWSCSAISSLPDVFITINGVQYVPPPSAYTLQSEGSCISGFGMNLPTESGELWI 300
 Db 303 SDGEMVWSCSAISSLPDVFITINGVQYVPPPSAYTLQSEGSCISGFGMDVPTESGELWI 362
 Qy 301 LGDVFIRQYFTVFDNRANNOVGLAPVA 326
 Db 363 LGDVFIRQYFTVFDNRANNOVGLAPVA 388

RESULT 6

PEMQAR
 N:pepsin A (EC 3.4.23.1) precursor - rhesus macaque
 N:Alternate names: pepsinogen A
 C:Species: Macaca mulatta (rhesus macaque)
 C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 18-Jun-1999
 C:Accession: J070309
 F:Evers, M.P.J.; Zelle, B.; Bebelman, J.P.; Pronk, J.C.; Mager, W.H.; Planta, R.J.; Erik
 Gene 65, 179-185, 1988
 A:Title: Cloning and sequencing of rhesus monkey pepsinogen A cDNA.
 A:Reference number: J070309; MUID:88313666
 A:Accession: J070309
 A:Molecule type: mRNA
 A:Residues: 1-388 <EN>
 A:Cross-references: GB:M20788; NID:g342274; PIDN:AAA36902.1; PID:g342275
 C:Superfamily: pepsin
 C:Keywords: aspartic proteinase; gastric juice; hydrolase; phosphoprotein; protein digest
 F:1-15/Domain: signal sequence #status predicted <SIG>
 F:16-388/Product: pepsinogen #status predicted <APT>
 F:63-388/Product: pepsinogen #status predicted <ENZ>
 F:94,277/Active site: Asp #status predicted
 F:107-112,268-272,311-344/Disulfide bonds: #status predicted
 F:130/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 96.4%; Score 1646; DB 1; Length 388;
 Best Local Similarity 95.1%; Pred. No. 2e-110;
 Matches 310; Conservative 13; Mismatches 3; Indels 0; Gaps 0;

Qy 1 VDEQPLENLDMEYFGTIGTIGTTPAQDFTVFDTSNNLWVPSVYCSLACTNHNRFNPD 60
 Db 63 IDEQPLENLDVEYFGTIGTIGTTPAQDFTVFDTSNNLWVPSVYCSLACTNHNLFNPD 122
 Qy 61 SSTQSTSETSVITYGTGSMTGILGYDTPVQGGISDTNQIFGLSETEPGSFLYYAPFDGI 120
 Db 123 SSTQSTSETSVITYGTGSMTGILGYDTPVQGGISDTNQIFGLSETEPGSFLYYAPFDGI 182

Qy 121 LGLAYPSISSSGATPVFDNIWQGLVSODLFSVYLSADDDQSGSVVIFGGIDSSYYTGS LN 180
 Db 183 LGLAYPSISSSGATPVFDNIWQGLVSODLFSVYLSADDDQSGSVVIFGGIDSSYYTGS LN 242
 Qy 181 WVPVTVEGYWQITVDSITMNGEATACAGCOAIVDTGTSLLTGTPSPANTQSDIGASEN 240
 Db 243 WVPVSVEGYWQISVDSITMNGEATACAGCOAIVDTGTSLLTGTPSPANTQSDIGASEN 302
 Qy 241 SDGDMVWSCSAISSLPDVFITINGVQYVPPPSAYTLQSEGSCISGFGMNLPTESGELWI 300
 Db 303 SDGEMVWSCSAISSLPDVFITINGVQYVPPPSAYTLQSEGSCISGFGMDVPTESGELWI 362
 Qy 301 LGDVFIRQYFTVFDNRANNOVGLAPVA 326
 Db 363 LGDVFIRQYFTVFDNRANNOVGLAPVA 388

RESULT 7

S19682
 N:pepsin A (EC 3.4.23.1) 4 precursor - Japanese macaque
 N:Alternate names: pepsinogen A isozyme 4
 C:Species: Macaca fuscata (Japanese macaque)
 C:Date: 22-Nov-1993 #sequence_revision 19-Oct-1995 #text_change 18-Jun-1999
 C:Accession: S19682; S16065
 R:Kageyama, T.; Tanabe, K.; Koiwai, O.
 Eur J. Biochem. 202, 205-215, 1991
 A:Title: Development-dependent expression of isozymogens of monkey pepsinogens and st
 A:Reference number: S19681; MUID:92037645
 A:Accession: S19682
 A:Molecule type: mRNA
 A:Residues: 1-388 <KAG>
 A:Cross-references: EMBL:X59753; NID:g38070; PIDN:CAA42425.1; PID:g38071
 A:Note: parts of sequence, including amino ends of pepsinogen and activation intermed
 C:Comment: This is a minor component of pepsin at all post-partum stages.
 C:Comment: Although two-step activation is observed, activation is predominantly a o
 C:Superfamily: pepsin
 C:Keywords: aspartic proteinase; gastric juice; hydrolase; phosphoprotein; protein di
 F:1-15/Domain: signal sequence #status predicted <SIG>
 F:16-388/Product: pepsinogen A 4 #status experimental <APT>
 F:16-62/Domain: activation peptide #status experimental <APT>
 F:63-388/Product: pepsin A 4 #status experimental <ENZ>
 F:38-39/Cleavage site: Leu-Iys (pepsin) #status experimental
 F:62-63/Cleavage site: Leu-Ile (pepsin) #status experimental
 F:94,277/Active site: Asp #status predicted
 F:107-112,268-272,311-344/Disulfide bonds: #status predicted
 F:130/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 93.4%; Score 1595; DB 1; Length 388;
 Best Local Similarity 92.0%; Pred. No. 8.8e-107;
 Matches 300; Conservative 19; Mismatches 7; Indels 0; Gaps 0;

Qy 1 VDEQPLENLDMEYFGTIGTIGTTPAQDFTVFDTSNNLWVPSVYCSLACTNHNRFNPD 60
 Db 63 IDEQPLENLDVEYFGTIGTIGTTPAQDFTVFDTSNNLWVPSVYCSLACTNHNLFNPD 122
 Qy 61 SSTQSTSETSVITYGTGSMTGILGYDTPVQGGISDTNQIFGLSETEPGSFLYYAPFDGI 120
 Db 123 SSTQSTSETSVITYGTGSMTGILGYDTPVQGGISDTNQIFGLSETEPGSFLYYAPFDGI 182
 Qy 121 LGLAYPSISSSGATPVFDNIWQGLVSODLFSVYLSADDDQSGSVVIFGGIDSSYYTGS LN 180
 Db 183 LGLAYPSISSSGATPVFDNIWQGLVSODLFSVYLSADDDQSGSVVIFGGIDSSYYTGS LN 242
 Qy 181 WVPVTVEGYWQITVDSITMNGEATACAGCOAIVDTGTSLLTGTPSPANTQSDIGASEN 240
 Db 243 WVPVSVEGYWQISVDSITMNGEATACAGCOAIVDTGTSLLTGTPSPANTQSDIGASEN 302
 Qy 241 SDGDMVWSCSAISSLPDVFITINGVQYVPPPSAYTLQSEGSCISGFGMNLPTESGELWI 300
 Db 303 SDGEMVWSCSAISSLPDVFITINGVQYVPPPSAYTLQSEGSCISGFGMDVPTESGELWI 362

QY 301 LGDVFIRQYFTVDRANNOVGLAPVA 326
|||||
Db 363 LGDVFIRQYFTVDRANNOVGLAPVA 388

RESULT 8

JC7245
pepsinogen A - common marmoset
C:Species: Callithrix jacchus (common marmoset)
C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 21-Jul-2000
C:Accession: JC7245
R:Kageyama, T.
J. Biochem. 127, 761-770, 2000
A:Title: New world monkey pepsinogens A and C, and prochymosins. Purification, character
A:Reference number: JC7245
A:Accession: JC7245
A:Molecule type: mRNA
A:Residues: 1-387 <KAG>
A:Cross-references: DDBJ:AB038384
A:Experimental source: strain NW794
C:Comment: This protein, a zymogen of pepsins, is the major proteolytic enzyme in verte
volution of mammalian orders and families.
C:Superfamily: pepsin
C:Keywords: gastric juice; zymogen

Query Match 91.2%; Score 1558; DB 2; Length 387;
Best Local Similarity 89.3%; Pred. No. 3.8e-104; -
Matches 291; Conservative 20; Mismatches 15; Indels 0; Gaps 0;

QY 1 VDEOPLNLYDMEYFGTIGTIGTPAQDFVVDGSSNLWVPSVYVCSLLACNHNHNPED 60
Db 62 IANOPLVNLYDMEYFGTIGTIGTPAQDFVVDGSSNLWVPSVYVCSLLACNHNHNPED 121
QY 61 SSTYQSTSETSYITVYTGSMGILGYDVTQVQVGGISDTNQIFGLSETEPGSFLYVAPFDGI 120
Db 122 SSTYQATSQTLIAVCTSMGILGYDVTQVQVGGIADTNQIFGLSETEPGSFLYVAPFDGI 181
QY 121 LGLAYPSISGGATPVFONINQGLVSQDLFSVYLSADDSQGVVIFGIDSSVYTGSLN 180
Db 182 LGLAYPSISGGATPVFONINQGLVSQDLFSVYLSADDSQGVVIFGIDSSVYTGSLN 241
QY 181 WYPTVEGYWQITVDSITMNGEAIACAGCAIVDTGTSLLTGPTSPANTQSDIGASEN 240
Db 242 WYPSAEGYWQITVDSITMNGEAIACAGCAIVDTGTSLLTGPTSPANTQSDIGASEN 301
QY 241 SDGDMVWSCATSSLPDVTFTINGVQVPPPSAYILOSEGSCISGFGQNNLPTESGELWI 300
Db 302 SNGEMVWSCATSSLPDVTFTINGVQVPPPSAYILOSEGSCISGFGQNNLPTESGELWI 361
QY 301 LGDVFIRQYFTVDRANNOVGLAPVA 326
Db 362 LGDVFIRQYFAVDRANNOVGLAPVA 387

RESULT 9

PEPG
pepsin A (EC 3.4.23.1) precursor - pig
N:Alternate names: pepsinogen A
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 24-Apr-1984 #sequence_revision 22-Jul-1994 #text_change 18-Jun-1999
C:Accession: J70307; A32455; B22434; A91410; A90185; A92039; A92179; PNO145; A00982
R:Tsuakagoshi, N.; Ando, Y.; Tomita, Y.; Uchida, R.; Takemura, T.; Sasaki, T.; Yamagata,
Gene 65, 285-292, 1988
A:Title: Nucleotide sequence and expression in Escherichia coli of cDNA of swine pepsin
al protein.
A:Reference number: J70307; MUID:88313677
A:Accession: J70307
A:Molecule type: mRNA
A:Residues: 1-386 <TSU>
A:Cross-references: GB:M20920; NID:g164601; PIDN:AAA31095.1; PID:g164602
R:Lin, X.; Wong, R.N.S.; Tang, J.
J. Biol. Chem. 264, 4482-4489, 1989

Query Match 88.3%; Score 1508.5; DB 1; Length 386;

A:Title: Synthesis, purification, and active site mutagenesis of recombinant porcine
A:Reference number: A32455; MUID:89174702
A:Accession: A32455
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-288,290-301,'Y',303-386 <LIN>
A:Cross-references: GB:J04601; NID:g164603; PIDN:AAA31096.1; PID:g164604
A:Note: Replacement of 91-Asp by Ala prevents autocatalytic activation
R:Ichihara, Y.; Sogawa, K.; Takahashi, K.
J. Biochem. 98, 483-492, 1985
A:Title: Isolation of human, swine, and rat prepepsinogens and calf preprochymosin, a
A:Reference number: A22434; MUID:86059312
A:Accession: B22434
A:Molecule type: protein
A:Residues: 1-19,'XX',22,'X',24-26 <ICH>
R:Moravsek, L.; Kostka, V.
FEBS Lett. 43, 207-211, 1974
A:Title: Complete amino acid sequence of hog pepsin.
A:Reference number: A91410; MUID:74299591
A:Accession: A91410
A:Molecule type: protein
A:Residues: 60-288,290-386 <MOR>
R:Stepanov, V.M.; Baratova, L.A.; Pugacheva, I.B.; Belyanova, L.P.; Revina, L.P.; Tim
Biochem. Biophys. Res. Commun. 54, 1164-1170, 1973
A:Title: N-terminal sequence of swine pepsinogen and pepsin. The site of pepsinogen a
A:Reference number: A90185; MUID:74031413
A:Accession: A90185
A:Molecule type: protein
A:Residues: 16-33,'D',35-118,'SD',121-127,'E',129-134 <STE>
A:Note: The authors point out the similarity of residues 31-43 and 59-71
R:Ong, E.B.; Perlmann, G.E.
J. Biol. Chem. 243, 6104-6109, 1968
A:Title: The amino-terminal sequence of porcine pepsinogen.
A:Reference number: A92039; MUID:69054241
A:Accession: A92039
A:Molecule type: protein
A:Residues: 16-33,'D',35-54,'AE' <ONG>
R:Sepulveda, P.; Marciniak, J.; Liu, D.; Tang, J.
J. Biol. Chem. 250, 5082-5088, 1975
A:Title: Primary structure of porcine pepsin. III. Amino acid sequence of a cyanogen
A:Reference number: A92179; MUID:75211282
A:Contents: active site
A:Accession: A92179
A:Molecule type: protein
A:Residues: 58-288,290-322,'D',324-349 <SEP>
A:Note: this is the final paper in a series
A:Note: variants having 314-Gln or an Ile between residues 288 and 290 were also found
R:Revina, L.P.; Vakhitova, E.A.; Pugacheva, I.B.; Lapuk, V.I.; Stepanov, V.M.
Biokhimiya 37, 1074-1080, 1972
A:Title: Investigation of peptides produced from N-terminal fragment of pepsin by hyd
A:Reference number: PNO145; MUID:73048197
A:Accession: PNO145
A:Molecule type: protein
A:Residues: 60-75,102-118,'SD',121-127,'E',129-134 <REV>
A:Note: article in Russian with English abstract
R:Andreeva, N.S.; Guschina, A.E.; Fedorov, A.A.; Shutzkever, N.E.; Volnova, T.V.
Adv. Exp. Med. Biol. 95, 23-31, 1977
A:Title: X-ray crystallographic studies of pepsin.
A:Reference number: A90016; MUID:78077917
A:Contents: annotation: X-ray crystallography, 3.0-2.7 angstroms
C:Comment: Minor amounts of the active enzyme occur with Ala-58 at the amino end.
C:Superfamily: pepsin
C:Keywords: aspartic proteinase; gastric juice; hydrolase; phosphoprotein; protein di
F:16-59/Domain: signal sequence #status experimental <SIG>
F:1-15/Domain: activation peptide #status experimental <APD>
F:60-386/Product: pepsin A #status experimental <MAT>
F:59-60/Cleavage site: Leu-Ile (pepsinogen) #status experimental
F:91,274/Active site: Asp #status experimental
F:104-109,265-269,309-342/Disulfide bonds: #status experimental
F:127/Binding site: phosphate (Ser) (covalent) #status experimental

Best Local Similarity 85.3%; Pred. No. 1.3e-100;
Matches 279; Conservative 33; Mismatches 14; Indels 1; Gaps 1;

Qy 1 VDEQPLENLDMEYFGTIGTIGTPAQDFTVFDGSSNLWVPSVYCSLACTNHNRFNPD 60
Db 60 IGDEPLENLDMEYFGTIGTIGTPAQDFTVFDGSSNLWVPSVYCSLACTNHNRFNPD 119
Qy 61 SSTQSTSETVITYGTGSMGTGILGYDVTQVQGGISDTNQIFGLSETEPGSFLLYAPFDGI 120
Db 120 SSTFEATSOELSIYGTGSMGTGILGYDVTQVQGGISDTNQIFGLSETEPGSFLLYAPFDGI 179
Qy 121 LGLAYPSISSGATPVFDNIWNGVLVSODLFSVYLSADDSQSVVIFGIDSSYYTGSIN 180
Db 180 LGLAYPSISASGATPVFDNLMDQGLVSODLFSVYLSNDDSGSVLLGIDSSYYTGSIN 239
Qy 181 WVPVTEGYWQITVDSITMNGEAIACAGCCQAIYDGTGSLTGPSTPIA-NIQSDIGASE 239
Db 240 WVPVTEGYWQITVDSITMNGEAIACAGCCQAIYDGTGSLTGPSTPIA-NIQSDIGASE 299
Qy 240 NSDGMVYSCSAISLPIVFTINGVQVPPSPAYILOEGSCISGFGMNLPTESGELW 299
Db 300 NSDGMVYSCSAISLPIVFTINGVQVPPSPAYILOEGSCISGFGMNLPTESGELW 359
Qy 300 ILGDVFIROYFTVFDNRANNQVGLAPVA 326
Db 360 ILGDVFIROYFTVFDNRANNQVGLAPVA 386

RESULT 10
E38302
pepsin (EC 3.4.23.-) IIIs precursor - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 14-Jun-1991 #sequence_revision 20-Sep-1991 #text_change 22-Jun-1999
C:Accession: E38302
R:Kageyama, T.; Tanabe, K.; Koiwai, O.
J. Biol. Chem. 265, 17031-17038, 1990
A:Title: Structure and development of rabbit pepsinogens. Stage-specific zymogens, nucle
A:Reference number: A38302; MUID:91009127
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-387 <KAG>
A:Cross-references: GB:M59237; GB:J05639; NID:g165597; PIDN:AAA85370.1; PID:g165598
C:Superfamily: pepsin
C:Keywords: aspartic proteinase; hydrolase; phosphoprotein; protein digestion

Query Match 85.1%; Score 1453; DB 2; Length 387;
Best Local Similarity 82.8%; Pred. No. 1.2e-96;
Matches 270; Conservative 32; Mismatches 24; Indels 0; Gaps 0;

Qy 1 VDEQPLENLDMEYFGTIGTIGTPAQDFTVFDGSSNLWVPSVYCSLACTNHNRFNPD 60
Db 62 VPTETLENLDTEFGTIGTIGTPAQDFTVFDGSSNLWVPSVYCSLACTNHNRFNPD 121
Qy 61 SSTQSTSETVITYGTGSMGTGILGYDVTQVQGGISDTNQIFGLSETEPGSFLLYAPFDGI 120
Db 122 SSTFOATSELSITYGTGSMGTGILGYDVTQVQGGISDTNQIFGLSETEPGSFLLYAPFDGI 181
Qy 121 LGLAYPSISSGATPVFDNIWNGVLVSODLFSVYLSADDSQSVVIFGIDSSYYTGSIN 180
Db 182 LGLAYPSISSSDATPVFDNMNNEGLVSODLFSVYLSDDSGSVLMFGIDSSYYTGSIN 241
Qy 181 WVPVTEGYWQITVDSITMNGEAIACAGCCQAIYDGTGSLTGPSTPIANTQSDIGASEN 240
Db 242 WVPVTEGYWQITVDSITMNGEAIACAGCCQAIYDGTGSLTGPSTPIANTQSDIGASEN 301
Qy 241 SDGMVYSCSAISLPIVFTINGVQVPPSPAYILOEGSCISGFGMNLPTESGELWI 300
Db 302 SDGMVYSCSAISLPIVFTINGVQVPPSPAYILOEGSCISGFGMNLPTESGELWI 361
Qy 301 LGDVFIRQYFTVFDNRANNQVGLAPVA 326

Query Match 85.1%; Score 1453; DB 2; Length 387;
Best Local Similarity 82.8%; Pred. No. 1.2e-96;
Matches 270; Conservative 32; Mismatches 24; Indels 0; Gaps 0;

Qy 1 VDEQPLENLDMEYFGTIGTIGTPAQDFTVFDGSSNLWVPSVYCSLACTNHNRFNPD 60
Db 62 VPTETLENLDTEFGTIGTIGTPAQDFTVFDGSSNLWVPSVYCSLACTNHNRFNPD 121
Qy 61 SSTQSTSETVITYGTGSMGTGILGYDVTQVQGGISDTNQIFGLSETEPGSFLLYAPFDGI 120
Db 122 SSTFOATSELSITYGTGSMGTGILGYDVTQVQGGISDTNQIFGLSETEPGSFLLYAPFDGI 181
Qy 121 LGLAYPSISSGATPVFDNIWNGVLVSODLFSVYLSADDSQSVVIFGIDSSYYTGSIN 180
Db 182 LGLAYPSISSSDATPVFDNMNNEGLVSODLFSVYLSDDSGSVLMFGIDSSYYTGSIN 241
Qy 181 WVPVTEGYWQITVDSITMNGEAIACAGCCQAIYDGTGSLTGPSTPIANTQSDIGASEN 240
Db 242 WVPVTEGYWQITVDSITMNGEAIACAGCCQAIYDGTGSLTGPSTPIANTQSDIGASEN 301
Qy 241 SDGMVYSCSAISLPIVFTINGVQVPPSPAYILOEGSCISGFGMNLPTESGELWI 300
Db 302 SDGMVYSCSAISLPIVFTINGVQVPPSPAYILOEGSCISGFGMNLPTESGELWI 361
Qy 301 LGDVFIRQYFTVFDNRANNQVGLAPVA 326

Db 362 LGDVFIRQYFTVFDNRANNQVGLAAAA 387

RESULT 11
D38302
pepsin (EC 3.4.23.-) II-4 precursor - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 14-Jun-1991 #sequence_revision 20-Sep-1991 #text_change 23-Feb-1997
C:Accession: D38302
R:Kageyama, T.; Tanabe, K.; Koiwai, O.
J. Biol. Chem. 265, 17031-17038, 1990
A:Title: Structure and development of rabbit pepsinogens. Stage-specific zymogens, nu
A:Reference number: A38302; MUID:91009127
A:Accession: D38302
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-387 <KAG>
A:Cross-references: GB:M59235; GB:J05638
C:Superfamily: pepsin
C:Keywords: aspartic proteinase; hydrolase; phosphoprotein; protein digestion

Query Match 80.9%; Score 1382; DB 2; Length 387;
Best Local Similarity 79.6%; Pred. No. 1.5e-91;
Matches 257; Conservative 30; Mismatches 36; Indels 0; Gaps 0;

Qy 1 VDEQPLENLDMEYFGTIGTIGTPAQDFTVFDGSSNLWVPSVYCSLACTNHNRFNPD 60
Db 62 VSTESLENLDAYFGTIGTIGTPQDFTVFDGSSNLWVPSVYCSLACTNHNRFNPD 121
Qy 61 SSTQSTSETVITYGTGSMGTGILGYDVTQVQGGISDTNQIFGLSETEPGSFLLYAPFDGI 120
Db 122 SSTQSTSETVITYGTGSMGTGILGYDVTQVQGGISDTNQIFGLSETEPGSFLLYAPFDGI 181
Qy 121 LGLAYPSISSGATPVFDNIWNGVLVSODLFSVYLSADDSQSVVIFGIDSSYYTGSIN 180
Db 182 LGLAYPSISSSDATPVFDNMNNEGLVSODLFSVYLSDDSGSVLMFGIDSSYYTGSIN 241
Qy 181 WVPVTEGYWQITVDSITMNGEAIACAGCCQAIYDGTGSLTGPSTPIANTQSDIGASEN 240
Db 242 WVPVTEGYWQITVDSITMNGEAIACAGCCQAIYDGTGSLTGPSTPIANTQSDIGASEN 301
Qy 241 SDGMVYSCSAISLPIVFTINGVQVPPSPAYILOEGSCISGFGMNLPTESGELWI 300
Db 302 LGENVYSCSAISLPIVFTINGVQVPPSPAYILOEGSCISGFGMNLPTESGELWI 361
Qy 301 LGDVFIRQYFTVFDNRANNQVGLA 323
Db 362 LGDVFIRQYFTVFDNRANNQVGLA 384

RESULT 12
C38302
pepsin (EC 3.4.23.-) II-2/3 precursor - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 14-Jun-1991 #sequence_revision 14-Jun-1991 #text_change 23-Feb-1997
C:Accession: C38302
R:Kageyama, T.; Tanabe, K.; Koiwai, O.
J. Biol. Chem. 265, 17031-17038, 1990
A:Title: Structure and development of rabbit pepsinogens. Stage-specific zymogens, nu
A:Reference number: A38302; MUID:91009127
A:Accession: C38302
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-387 <KAG>
A:Cross-references: GB:J05638
C:Superfamily: pepsin
C:Keywords: aspartic proteinase; hydrolase; phosphoprotein; protein digestion

Query Match 80.4%; Score 1374; DB 2; Length 387;
Best Local Similarity 78.9%; Pred. No. 5.5e-91;
Matches 255; Conservative 31; Mismatches 37; Indels 0; Gaps 0;

```
QY 1 VDEQPLENYLDMYEGTIGTIGTTPAQDFTVVFDTGSSNLWVPVSVYCSSLACTNHNRFNPED 60
| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 62 VSTESMENYLDAYEGTISIGTTPQDFTVFDTGSSNLWVPVSVYCSSLACTNHNRFNPED 121
| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 STYQSTSTVSTVYTGTSMTGILGYDVTVOVGGISDTNQIFGLSETEPGSFYLYYAPFDGI 120
| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 122 STYQSTSTVSTVYTGTSMTGILGYDVTVOVGGISDTNQIFGLSETEPGSFYLYYAPFDGI 181
| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 LGLAYPSISSSGATPVFDNIWNOGLVSDQLFSVYLSADDSQGSVVFIFGIDSSYTGSLN 180
| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 182 LGLAYPSISSSGATPVFDNIWNOGLVSDQLFSVYLSADDSQGSVVFIFGIDSSYTGSLN 241
| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 181 WYPVTVGHWQITVDSITMNGBAICAECAEQQAIVDTGTSLLTGPTSPITANIOSDTCASEN 240
| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 242 WYPVTVGHWQITVDSITMNGBAICAECAEQQAIVDTGTSLLTGPTSPITANIOSDTCASEN 301
| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 241 SGDDMVVSCAISLSPDIVFTINGVQVPPPSAYILQSEGCISGFGQGNLPTSEGLWI 300
| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 302 LGGENVISCASLSPDIVFTINGVQVPPPSAYILQSEGCISGFGQGNLPTSEGLWI 361
| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 301 LGDVFIRQYFTVFDNRANNOVGLA 323
| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 362 LGDVFIRQYFTVFDNRANNOVGLA 384
| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 13
B38302
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 14-Jun-1991 #sequence_revision 20-Sep-1991 #text_change 23-Feb-1997
C:Accession: B38302
R:Kageyama, T.; Tanabe, K.; Koiwai, O.
J. Biol. Chem. 265, 17031-17038, 1990
A:Title: Structure and development of rabbit pepsinogens. Stage-specific zymogens, nucle
A:Reference number: A38302; MUID:91009127
A:Accession: B38302
A>Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-387 <KAG>
A:Cross-references: GB:M59235; GB:J05638
C:Superfamily: pepsin
C:Keywords: aspartic proteinase; hydrolase; phosphoprotein; protein digestion

Query Match 78.9%; Score 1348; DB 2; Length 387;
Best Local Similarity 77.3%; Pred. No. 3.9e-89;
Matches 252; Conservative 32; Mismatches 42; Indels 0; Gaps 0;

QY 1 VDEQPLENYLDMYEGTIGTIGTTPAQDFTVVFDTGSSNLWVPVSVYCSSLACTNHNRFNPED 60
| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 62 VSTESMENYLDAYEGTISIGTTPQDFTVFDTGSSNLWVPVSVYCSSLACTNHNRFNPED 121
| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 STYQSTSTVSTVYTGTSMTGILGYDVTVOVGGISDTNQIFGLSETEPGSFYLYYAPFDGI 120
| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 122 STYQSTSTVSTVYTGTSMTGILGYDVTVOVGGISDTNQIFGLSETEPGSFYLYYAPFDGI 181
| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 LGLAYPSISSSGATPVFDNIWNOGLVSDQLFSVYLSADDSQGSVVFIFGIDSSYTGSLN 180
| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 182 LGLAYPSISASATPVFDNIWNOGLVSDQLFSVYLSADDSQGSVVFIFGIDSSYTGSLN 241
| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 181 WYPVTVGHWQITVDSITMNGBAICAECAEQQAIVDTGTSLLTGPTSPITANIOSDTCASEN 240
| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 242 WYPVTVGHWQITVDSITMNGBAICAECAEQQAIVDTGTSLLTGPTSPITANIOSDTCASEN 301
| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 241 SGDDMVVSCAISLSPDIVFTINGVQVPPPSAYILQSEGCISGFGQGNLPTSEGLWI 300
| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 302 LGGENVISCASLSPDIVFTINGVQVPPPSAYILQSEGCISGFGQGNLPTSEGLWI 361
| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 301 LGDVFIRQYFTVFDNRANNOVGLAPVA 326
| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 362 LGDVFIRQYFTVFDNRANNOVGLAANA 387
| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
RESULT 14
JC7574
pepsinogen A - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 31-Mar-2001 #sequence_revision 31-Mar-2001 #text_change 31-Mar-2001
C:Accession: JC7574; PC7119
R:Ikuzawa, M.; Inokuchi, T.; Kobayashi, K.; Yasumasu, S.
J. Biochem. 129, 147-153, 2001
A:Title: Amphibian pepsinogens: Purification and characterization of Xenopus pepsinog
A:Reference number: JC7573; PMID:11134969
A:Contents: Stomach
A:Accession: JC7574
A:Molecule type: mRNA
A:Residues: 1-384 <IKU>
A:Cross-references: DDBJ:AB045380
A:Accession: PC7119
A:Residues: 16-35;57-76 <IK2>
C:Comment: This protein is a zymogen for gastric aspartic proteinase, with pepsin-lik
C:Genetics:
A:Gene: Pga
C:Keywords: stomach; zymogen

Query Match 76.8%; Score 1311; DB 3; Length 384;
Best Local Similarity 71.8%; Pred. No. 1.7e-86;
Matches 232; Conservative 57; Mismatches 34; Indels 0; Gaps 0;

QY 4 OPLENYLDMYEGTIGTIGTTPAQDFTVVFDTGSSNLWVPVSVYCSSLACTNHNRFNPEDSST 63
| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 62 ETLQNYMDIEYGTISIGTTPQDFTVFDTGSSNLWVPVSVYCSSLACTNHNRFNPEDSST 121
| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 64 YQSTSTVSTVYTGTSMTGILGYDVTVOVGGISDTNQIFGLSETEPGSFYLYYAPFDGILGL 123
| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 122 FOATNTPVSIQYTGSMGFLGYDVLQVGNIOISNQMFGLSESEPGSFYLYYAPFDGILGL 181
| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 124 AYPSTSSSGATPVFDNIWNOGLVSDQLFSVYLSADDSQGSVVFIFGIDSSYTGSLNWVP 183
| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 182 APPSIASSQATPVFDNIWNOGLVSDQLFSVYLSADDSQGSVVFIFGIDSSYTGSLNWVP 241
| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 184 VIVEGYWQITVDSITMNGBAICAECAEQQAIVDTGTSLLTGPTSPITANIOSDTCASENSG 243
| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 242 LTAETWQITVDSITMNGBAICAECAEQQAIVDTGTSLLTGPTSPITANIOSDTCASENSG 301
| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 244 DMVVCSAISLSPDIVFTINGVQVPPPSAYILQSEGCISGFGQGNLPTSEGLWILGD 303
| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 302 QVINCNNISNPTIVFTINGVQVPLSPSAYVYRQNGCGSSGFGQGNLPTSEGLWILGD 361
| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 304 VFIRQYFTVFDNRANNOVGLAPVA 326
| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 362 VFIRQYFTVFDNRANNOVGLAPVA 384
| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15
JC7575
pepsinogen A - bullfrog
C:Species: Rana catesbeiana (bullfrog)
C:Date: 31-Mar-2001 #sequence_revision 31-Mar-2001 #text_change 31-Mar-2001
C:Accession: JC7575
R:Ikuzawa, M.; Inokuchi, T.; Kobayashi, K.; Yasumasu, S.
J. Biochem. 129, 147-153, 2001
A:Title: Amphibian pepsinogens: Purification and characterization of Xenopus pepsinog
A:Reference number: JC7573; PMID:11134969
A:Contents: Stomach
A:Accession: JC7575
A:Molecule type: mRNA
A:Residues: 1-385 <IKU>
A:Cross-references: DDBJ:AB045376
C:Comment: This protein is a zymogen for gastric aspartic proteinase, with pepsin-lik
C:Genetics:
A:Gene: Pga
C:Keywords: stomach; zymogen
```

Search completed: September 6, 2001, 16:45:57
Job time: 501 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:39:40 ; Search time 113.12 Seconds
(without alignments)
59.339 Million cell updates/sec

Title: US-09-603-713-31

Perfect score: 1708
Sequence: 1 VDEQPLENYLDMEYFGTGT.....RQYTFVFDNRANNOVGLAPVA 326

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA: *

database ;
issued_patents_an.
1: /cqn2_6/ptodata/2/laa/5A COMB.pcp:*

2: /cgn2_6/ptodata/2/laa/5B_COMB.pep.*

3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*

```
4: /cgn2_6/ptodata/2/laa/6B_COMB.per:*
5: /cgn2_6/ptodata/2/laa/6B_COMB.COMB
```

```
3: /cgn2_6/ptodata/2/1aa/PC1US_COMB.pe
6: /cgn2_6/ptodata/2/1aa/BACKFILES.pe
```

U: /cgnz_v/prouctn/z/rnu/mucnrrrresr.pel

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	1008.5	59.0	427	2	US-08-846-021A-8	Sequence 8, Appli
2	988.5	57.9	396	1	US-08-208-007A-13	Sequence 13, Appli
3	988.5	57.9	396	4	US-09-032-523-9	Sequence 9, Appli
4	981	57.4	458	6	5217891-15	Patent No. 5217891
5	825	48.3	349	4	US-09-032-523-3	Sequence 3, Appli
6	822.5	48.2	412	1	US-08-208-007A-12	Sequence 12, Appli
7	822.5	48.2	412	4	US-08-974-691-4	Sequence 4, Appli
8	736	43.1	419	4	US-08-974-691-3	Sequence 3, Appli
9	729	42.7	420	4	US-09-008-271A-4	Sequence 4, Appli
10	729	42.7	420	4	US-08-974-691-8	Sequence 8, Appli
11	700.5	41.0	397	3	US-09-079-415-2	Sequence 2, Appli
12	694.5	40.7	398	1	US-08-328-314-2	Sequence 2, Appli
13	694.5	40.7	398	1	US-08-731-045-2	Sequence 2, Appli
14	668	39.1	445	4	US-08-974-691-6	Sequence 6, Appli
15	668	39.1	451	4	US-08-974-691-2	Sequence 2, Appli
16	658.5	38.6	409	1	US-08-360-673-6	Sequence 6, Appli
17	650	38.1	410	1	US-08-088-633-2	Sequence 2, Appli
18	650	38.1	410	1	US-08-245-756-2	Sequence 2, Appli
19	650	38.1	410	1	US-08-441-750-2	Sequence 2, Appli
20	650	38.1	410	2	US-08-441-751-2	Sequence 2, Appli
21	650	38.1	410	5	PCR-US92-02521-2	Sequence 2, Appli
22	648.5	38.0	395	1	US-08-723-938-3	Sequence 3, Appli
23	648.5	38.0	395	2	US-09-080-538-3	Sequence 3, Appli
24	439	25.7	430	1	US-08-535-237-2	Sequence 2, Appli
25	427.5	25.0	427	1	US-07-958-222A-2	Sequence 2, Appli
26	330	19.3	330	3	US-08-115-753-1	Sequence 1, Appli
27	330	19.3	419	3	US-08-115-753-2	Sequence 2, Appli

63 TYQSTSETVITYGTGSMITGILGYDTVQVGGISDITNOIFGLSETPEPGFLYVAPFGIILG 122
127 TYSQPGSFIQXGTGSLGIGADQVSVGLTVWQGFQGESVTEPGQTFVDAEFDGILG 186
123 LAYPSISSGATPVFONINWQGLVSQDLFSVYLSADDO--SGSVWIFGGIDSSYYTGSLSN 180
187 LGYPSLAVGAVTVFVFNMAQNLVDFMFVSVYSSNPPEGAGSELIFGGVDHSHFSGSLN 246
181 WVPVTEGVTQITVDSITMNGEAIACAEGCOAIVDTGTSLTGTPTPIANIOSDIGASEN 240
247 WVPVTKAYQWIALDNIOVGTVMFCEGCOAIVDTGTSLTGTSPDKIKOLQNAIGAAP- 305
241 SGGDMVVSQSAISLDPDIVFTINGVQVPPSPAYILQS--EGS--CISGFGQGNLPTESG 296
306 VDEYAVECANLVNMPDVTFTINGVPYTLSTPTATLLDFDVGQFCSSGFGGLDIHPPAG 365
297 ELWILGDFVFIROYFTVFDNRANNQVGLAP 324
366 PLWILGDFVFIROYFVFDNRNNRVGLAP 393

RESULT 3
US-09-032-523-9
; Sequence 9, Application US/09032523
; Patent No. 6232454
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl
; APPLICANT: Baugh, Mariah
; TITLE OF INVENTION: HUMAN PROTEINASE MOLECULES
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/032,523
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0479 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Genbank
; CLONE: 181994
; US-09-032-523-9

Query Match 57.9%; Score 988.5; DB 4; Length 396;

171 ONLGRPLSIHYGTGSMOILGYDTVTSNIVDIOQTGVLSTQEPEDFTVAEFDGILGMA 230
125 YPSISSGATPVFONINWQGLVSQDLFSVYLSADDOQSGSVVIFGGIDSSYYTGSLSNWPV 184
231 YPSLASEYVIPFVFNMAQNLVDFMFVSVYSSNPPEGAGSELIFGGVDHSHFSGSLNWPV 289
185 TVEGWQITVDSITMNGEAIACAEGCOAIVDTGTSLTGTPTPIANIOSDIGASENSDGD 244
290 TVQQVQWQITVDSITMNGEAIACAEGCOAIVDTGTSLTGTPTPIANIOSDIGASENSDGD 349
245 MVVVSQSAISLDPDIVFTINGVQVPPSPAYILQS--EGS--CISGFGQGNLPTESGELWILG 304
350 FIDICDNLSYVTVFVFNMAQNLVDFMFVSVYSSNPPEGAGSELIFGGVDHSHFSGSLN 405
305 FRIQYFTVFDNRANNQVGLA 323
406 FRIEYYSVFDNRANNVGLA 424

RESULT 2
US-08-208-007A-13
; Sequence 13, Application US/08208007A
; Patent No. 5501969
; GENERAL INFORMATION:
; APPLICANT: HASTINGS, ET AL.
; TITLE OF INVENTION: Human Osteoclast-Derived Cathepsin
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/208,007A
; FILING DATE: March 8, 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: No. 5501969e
; FILING DATE: No. 5501969e
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-95
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; US-08-208-007A-13

Query Match 57.9%; Score 988.5; DB 1; Length 396;
Best Local Similarity 57.0%; Pred. No. 3.2e-80;
Matches 187; Conservative 55; Mismatches 79; Indels 7; Gaps 4;

3 EOPLENLDMEYFTIGTIGTPADQFTVVDFTGSSNLWVPSVYSSNACTNHNFPEDSS 62
67 KEPLINLDMEYFTIGTIGTSPQNFVIFDTGSSNLWVPSVYSSNACTNHNFPEDSS 126

[illegible]

Db 122 KYNDSKSTYVKNKTSFDIHYGSGSLGYLSQDTSVSPQCSASSALGCVKVERQVGE 181
QY 104 SETEPGSLFYAPFGILGLAYPSISSSGATPVFDNIWNOGLVSDLFVYLSADD--QS 161
Db 182 ATKQPGITIAAKFGILGMAYPRISVNNVLPVFDNLMOOKLVQDNIFSYLSRDPDAQ 241
QY 162 GSVIFGIGDSSYYTGSLLMWVPTVEGYWQITVDSITMNGEAIACAEGCOAIVDTGTSLL 221
Db 242 GGELMLGGTDSKYKGLSYLVNTRKAYQVHLDQVEVASGLTLCKEGCEAIVDTGTSLL 301
QY 222 TGPTSPIANQSDIGASENSDGMVSCSAISSLPDIVFTINGVQVPPPSAYILO---- 277
Db 302 VGPDEVRELOAKAIGVPLVIOEYIPCKEYVSTLPAITLKGKGYKUSPEYTYTLKVSOA 361
QY 278 SEGSCISGFGMNLPTESGELWILGDFVFIROYFTVFDNRANNOVGLAPVA 326
Db 362 GKTCLSGPMGMDIPPPSGFLMGDLGVFCIGRYTTFVDRDNRNRVGAEEA 410
RESULT 8
US-08-974-691-3
; Sequence 3, Application US/08974691
; Patent No. 6225103
; GENERAL INFORMATION:
; APPLICANT: Keolsch, Gerald
; APPLICANT: Lin, Xinli
; APPLICANT: Tang, Jordan
; TITLE OF INVENTION: Cloning and Characterization of Napsin
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center, 1201 W. Peachtree
; STREET: St.
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/974,691
; FILING DATE: 20-NOV-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/031,196
; FILING DATE: 20-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/046,126
; FILING DATE: 09-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: OMRF 166
; TELEPHONE: 404-873-8794
; TELEFAX: 404-873-8795
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 419 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-974-691-3

Query Match 43.1%; Score 736; DB 4; Length 419;
Best Local Similarity 43.1%; Pred. No. 1,1e-57;
Matches 140; Conservative 60; Mismatches 117; Indels 8; Gaps 3;

QY 5 PLENLDMEYFCTIGTIGTQAQDFTTVVFDGSSNLWVPSVYCS--SLACTNHNRFNPEDESS 62
Db 64 PLKSKMNTQYFCTIGLTGTPPQNFVTVDGSSNLWVPSRCHFFSLACWFHHRFNPKASS 123
QY 63 TYQSTSETVSIYTCGSMFGILGYDTQVGGISDITNQIFGLSETEPGSLFYAPFGILG 122
Db 124 SRPNCTKFAIYGTCRSLGSLSDQNLITIGGIDHAFVTFGEALWEPSSLFALAHFDGILG 183
QY 123 LAYPSISSGATPVFDNIWNOGLVSDLFVYLSADDQ--SGSVVIFGIGDSSYYTGSLLN 180
Db 184 LGFPTLAVGCVOPPLDAMVEQGLLEKPEVSYFLNRDSESGDGLVGLGSDPAHYVPLT 243
QY 181 WYPTVEGYWQITVDSITMNGEAIACAEGCOAIVDTGTSLLTGPTSPIANQSDIGASEN 240
Db 244 FIPVT--PAYWQVHMESVKVGTGLSLCAQCSAILDTGTSLITGSPSEIRALNKAIGGYPF 303
QY 241 SGDDMVVSCSAISSLPDIVFTINGVQVPPPSAYILO----SEGSCISGFGMNLPTESG 296
Db 304 LNGQYFIQCSKTPTLPVVSFHLGGVWFNLTDQDYVIQDQLQSDVGLGCLLGLFQALDIPK 363
QY 297 ELWILGDFVFIROYFTVFDNRANNOVG 321
Db 364 PLWILGDFVLPYVAVFDRGDKNVG 388
RESULT 9
US-09-008-271A-4
; Sequence 4, Application US/09008271A
; Patent No. 6203979
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; Hillman, Jennifer L.
; Yue, Henry
; Guegler, Karl J.
; Corley, Neil C.
; Tang, Tom Y.
; Shah, Purvi
; TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/008,271A
; FILING DATE: 16-Jan-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: <Unknown>
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Mohan-Peterson, Sheela
; REGISTRATION NUMBER: 41,201
; REFERENCE/DOCKET NUMBER: PF-0458 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 420 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: LUNCAST01
; CLONE: 877617

SEQUENCE DESCRIPTION: SEQ ID NO: 4 ;
US-09-008-271A-4

Query Match 42.7%; Score 729; DB 4; Length 420;
Best Local Similarity 43.8%; Pred. No. 4.5e-57;
Matches 145; Conservative 60; Mismatches 114; Indels 12; Gaps 4;

```
QY 5 PLENLDMEYFGTIGTGPADQFTVFTGSSNLWVPSVYCS--SLACTNHNRPEDSS 62
DB 69 PLSNRYDQYFGEIGLTPPQNFVAFDTGSSNLWVPSRRCHFFSVPCWLHHRDPKASS 128
QY 63 TYQSTSEVSIYTGTSMTGILGYDTVOVGGISDTNQIFGLSETEPGSFLYAYAPDGIIG 122
DB 129 SFQANGTKFAIQYGRVGDILSEDKLIGGKASVIFGEALWEPSSLVFAFAHFDGILG 188
QY 123 LAYPSISSGATPVDNINWQGLSODLFSVYLSAD--DQSGSVVIFGIDSSYYTGSIN 180
DB 189 LGFPILSVGEVRPPMDVLVEQGLDKPVSEFLNRPDPEPDGGLVVGSDPAHYIPPLT 248
QY 181 WPVTVVEGWQITVDSITMNGEAIACAGCOAIVDTGTSLLTGPTSPANTQSDIGASEN 240
DB 249 FVPVTPAYQIHMERYKVGPGTLCAKGCACAILDTGTSLLTGPTSPANTQSDIGASEN 308
QY 241 SDGDMVSCAIISSLPDIVFTINGVQYVPVPSAYILQSEGS---CISGFGMNLPTESG 296
DB 309 LAGEYIILCSEIKLPVAVSFLGGWFNLTAHDYVIQITRNGVRLCSGQALDVPVPPAG 368
QY 297 ELWILGDVFIROYFTVDR---ANNQVGLA 323
DB 369 PFWILGDVFLGYVAVFDRGDMKSSARVGLA 399
```

RESULT 10

US-08-974-691-8
; Sequence 8, Application US/08974691
; Patent No. 6225103
; GENERAL INFORMATION:
; APPLICANT: Keolsch, Gerald
; APPLICANT: Lin, Xianli
; APPLICANT: Tang, Jordan
; TITLE OF INVENTION: Cloning and Characterization of Napsin
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center, 1201 W. Peachtree
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30109-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/974,691
; FILING DATE: 20-NOV-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/031,196
; FILING DATE: 20-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/046,126
; FILING DATE: 09-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: OMFR 166
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-873-8794
; TELEFAX: 404-873-8795

INFORMATION FOR SEQ ID NO: 8;

SEQUENCE CHARACTERISTICS:
; LENGTH: 420 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
US-08-974-691-8

Query Match 42.7%; Score 729; DB 4; Length 420;
Best Local Similarity 43.8%; Pred. No. 4.5e-57;
Matches 145; Conservative 60; Mismatches 114; Indels 12; Gaps 4;

```
QY 5 PLENLDMEYFGTIGTGPADQFTVFTGSSNLWVPSVYCS--SLACTNHNRPEDSS 62
DB 69 PLSNRYDQYFGEIGLTPPQNFVAFDTGSSNLWVPSRRCHFFSVPCWLHHRDPKASS 128
QY 63 TYQSTSEVSIYTGTSMTGILGYDTVOVGGISDTNQIFGLSETEPGSFLYAYAPDGIIG 122
DB 129 SFQANGTKFAIQYGRVGDILSEDKLIGGKASVIFGEALWEPSSLVFAFAHFDGILG 188
QY 123 LAYPSISSGATPVDNINWQGLSODLFSVYLSAD--DQSGSVVIFGIDSSYYTGSIN 180
DB 189 LGFPILSVGEVRPPMDVLVEQGLDKPVSEFLNRPDPEPDGGLVVGSDPAHYIPPLT 248
QY 181 WPVTVVEGWQITVDSITMNGEAIACAGCOAIVDTGTSLLTGPTSPANTQSDIGASEN 240
DB 249 FVPVTPAYQIHMERYKVGPGTLCAKGCACAILDTGTSLLTGPTSPANTQSDIGASEN 308
QY 241 SDGDMVSCAIISSLPDIVFTINGVQYVPVPSAYILQSEGS---CISGFGMNLPTESG 296
DB 309 LAGEYIILCSEIKLPVAVSFLGGWFNLTAHDYVIQITRNGVRLCSGQALDVPVPPAG 368
QY 297 ELWILGDVFIROYFTVDR---ANNQVGLA 323
DB 369 PFWILGDVFLGYVAVFDRGDMKSSARVGLA 399
```

RESULT 11

US-09-079-415-2
; Sequence 2, Application US/09079415
; Patent No. 6013452
; GENERAL INFORMATION:
; APPLICANT: Christensen, Tove
; APPLICANT: Lehmebeck, Jan
; TITLE OF INVENTION: A Fungus Wherein The area; pepC and/or
; TITLE OF INVENTION: pepe Genes Have Been Inactivated
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6013452o No. 6013452disk of No. 6013452th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/079,415
; FILING DATE: 14-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Rozek, Carol
; REGISTRATION NUMBER: 36,993
; REFERENCE/DOCKET NUMBER: 4657.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123

```
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 397 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-079-415-2

Query Match 41.0%; Score 700.5; DB 3; Length 397;
Best Local Similarity 41.9%; Pred. No. 1.4e-54;
Matches 134; Conservative 68; Mismatches 115; Indels 3; Gaps 2;

QY 6 LENYLDMEYFGIGTGAQDFTVFDGSSNLWVPSVYCSSLACTNNHNFNPDSSTYQ 65
DB 76 VDNFLNAQYFSEIEIGTPQKFKVLDTGSSNLWVPSSECGSIACVYLNKDYSSSTYQ 135
QY 66 STSETVSYTGSGMTGILGYDTVQVGGISDNTQIFGLSETPGSFYAPFDGILGLAY 125
DB 136 KNGSEFAIRYKSGSLGFGYSQDTLTKIGDLKVKDQLFAEATSEPLGAFAGFREDGILGLGF 195
QY 126 PSTSSSGATPVFDNIWNOGLVSDLSVYLSADDDSG--SVVIFGIDSSYVTGSLNWVP 183
DB 196 DTISVKNKIPPPYFSLMDQGLLDEPVFAFYLGDTNKGDDSVATFGVGDVNDHVTGELVKIP 255
QY 184 VVVEGYWQITVDSITMNGEAIACAEQCAIVDTGTSLLTGPTSPITANIQSDIGASENSDG 243
DB 256 LRRKAWVEVDLDAIAL-GDSVAELDWTGVLDTGSLIALATTLAELINKEIKGKGTG 314
QY 244 DMVVSCTASSLPDVIPTFTINGVOYVPPPSAYILQSGSCISGFGQGNLPTSEGELWILD 303
DB 315 QYSDVCDKRDLSPLDTFTLSGYNFTIGPYDTLEVOGSCISAFMGDMDFPVPVGLAILGD 374
QY 304 VFIROYETVFDNRANNOVGLA 323
DB 375 APLRKWTSYVDLANGAVGLA 394

RESULT 12
US-08-328-314-2
; Sequence 2, Application US/08328314
; Patent No. 5674728
; GENERAL INFORMATION:
; APPLICANT: Buxton, Frank
; APPLICANT: Jarai, Gabor
; APPLICANT: Visser, Jacob
; TITLE OF INVENTION: Fungal Protease
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII editor
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,314
; FILING DATE: TBA
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sprull, W. Murray
; REGISTRATION NUMBER: 32,943
; REFERENCE/DOCKET NUMBER: 4-19746/A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919) 541-8615
; TELEFAX: (919) 541-8689
; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 398 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-328-314-2

Query Match 40.7%; Score 694.5; DB 1; Length 398;
Best Local Similarity 40.9%; Pred. No. 4.9e-54;
Matches 137; Conservative 63; Mismatches 122; Indels 13; Gaps 3;

QY 1 VDEQPL-----ENYLDMEYFGTIGTGAQDFTVFDGSSNLWVPSVYCSSLAC 50
DB 62 VEENPINDMSRHDVLNDFLNAQYFSEIEIGTPQKFKVLDTGSSNLWVPSSECGSIAC 121
QY 51 THNRFNPEDSSTYQSTSETVSYTGSGMTGILGYDTVQVGGISDNTQIFGLSETEPGS 110
DB 122 YLHNKYDSSASTYHKNSEPAIKYSGSLGFSVSDTLKIGDLKVKGDFAEATNEPGL 181
QY 111 FLYYAPFDGILGLAYPSISSSGATPVFDNIWNOGLVSDLSVYLSADDDSG--SVVIFG 168
DB 182 AFAFGREFGILGLGYDTISVKNKIVPPFYNMLDQGLLDEPVFAFYLGDTNKGDESVA 241
QY 169 GIDSSYTGSLNWVPVTVEGYWOITVDSITMNGEAIACAEQCAIVDTGTSLLTGPTSP 228
DB 242 GVDKDHVTGELIKIPLRRKAYWEVELDAIAL-GDDVAEMENTGVILDTGTSIALPADLA 300
QY 229 ANIOSDIGASENSCDMVVSCSAISSLPDVIPTFTINGVOYVPPPSAYILQSGSCISG 288
DB 301 EMINAQIGAKKWTGQYTVDCDKRSSLPDVTFTLAGHNFTISSYDITLEVQSGSCVSA 360
QY 289 MNLPTSEGELWILDVFIROYETVFDNRANNOVGLA 323
DB 361 MDPPEPVGLAILGDFAFLRKWYSVYDLGNSAVGLA 395

RESULT 13
US-08-731-045-2
; Sequence 2, Application US/08731045
; Patent No. 5756338
; GENERAL INFORMATION:
; APPLICANT: Buxton, Frank
; APPLICANT: Jarai, Gabor
; APPLICANT: Visser, Jacob
; TITLE OF INVENTION: Fungal Protease
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY CORPORATION
; STREET: 520 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591-9725
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII Editor
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/731,045
; FILING DATE: Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/328,314
; FILING DATE: October 24, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferraro, Gregory D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 4-19746/A/DIV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 277-3318
; TELEFAX: (908) 277-4306
```

```

; INFORMATION FOR SEQ ID NO: 2:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 398 amino acids
;
; TYPE: amino acid
;
; TOPOLOGY: linear
;
; MOLECULE TYPE: protein
US-08-731-045-2

```

Query Match	40.7%;	Score 694.5;	DB 1;	Length 398;
Best Local Similarity	40.9%;	Pred. NO. 4.9e-54;		
Matches 137; Conservative	63;	Mismatches 122;	Indels 13;	Gaps 3;

QY	1	VDEOPL-----ENYLDMEYFGTIGTGPQDFTVVDGTGSSNLWVPSVCCSSLAC	50
Db	62	VEENPINDMRHVDVNDFNLAQYSEIELGTPOQKFVKVLDGTGSSNLWVPSSECSIAAC	121
QY	51	TNENRPEDSSYQSTSETSVITYGTSGMTGLGVDTVQVGGISDTNOIFGLSETPGS	110
Db	122	YLHNKYDSSASSYTHKNGSEFAIKYGGSLGFPVSDDTLKIGDLKYKGQDFAETNEPGL	181
QY	111	FLYYAPDGTGLGLAYSISSSGATPVFDNIWNOGLVSOQLFSVYLSADDSQG--SVWIFG	168
Db	182	AFAGFRDGLGLUGYDTISVKNKTVPPFYNNLDOGLLDEPFVAFYLGDTNKEGDES	241
QY	169	GIDSSYTYTGLSNVPVTVGTYOITVDSITTMGEALACAEQGQAIIVDTCTSLLTGPTSI	228
Db	242	GVDKDHWTGELIKIPURRAYVEVELDAIAL--GDDVAEMENTGVILDTCTSLIALPADLIA	300
QY	229	ANIQSDIGASENSGDWVYSCSAISLSPDIVTINGVQYVPVPSAYILOSEGSCTIGGQF	288
Db	301	EMINAQIGAKKWTGGQYTVDCOKRSLSPDVTTLAGHNFTISSDYTLVEQSCVSAFMG	360
QY	289	MNLPTESGELWILGDVFIROYTFVFDRAANNQVGLA	323
Db	361	MDPPEVGEFLAILGDFAFLRKWTSVYDLGNSAVGLA	395

RESULT 14
US-08-974-691-6
Sequence 6, Application US/08974691
Patent No. 6225103
GENERAL INFORMATION:
APPLICANT: Keolsch, Gerald
APPLICANT: Lin, Xinli
APPLICANT: Tang, Jordan
TITLE OF INVENTION: Cloning and Characterization of Napsin
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlanta Center, 1201 W. Peachtree
CITY: Atlanta
STATE: GA
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,691
FILING DATE: 20-NOV-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/031,196
FILING DATE: 20-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/046,126
FILING DATE: 09-MAY-1997
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.

```

: REGISTRATION NUMBER: 31,284
: REFERENCE/DOCKET NUMBER: OMR# 166
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 404-873-8794
: TELEFAX: 404-873-8795
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 445 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: ORIGINAL SOURCE:
: ORGANISM: Homo sapiens
: US-08-974-691-6

```

Query Match	39.1%	Score 668;	DB 4;	Length 445;
Best Local Similarity	41.1%;	Pred. No. 1.3e-51;		
Matches 136;	Conservative 60;	Mismatches 123;	Indels 12;	Gaps 5;

QY	5	PLENYLDMEYEGTIGICTPAQDFTVVFEDTGSNNLWPSVYCS--SLACTHNHNRPNPEDSS	62
Db	69	PLSKFLDAQYGEIGLGTTPQNFVAFDTCSSNLWPSRRCHFFSVPCWPFHFRNPASS	128
QY	63	TYOSTSETVSTIGTSGMTGILGVDTVQVGGISDINOICLSETEPCGSFLYYAPFGGILG	122
Db	129	SFRPSTGKFAIQYGTGRVDGILSEDKLTIGGIKCAVSIFGEALWESSLVTVSRPDGILG	188
QY	123	LAYPSTSSSGATPVFDNIWNOGLVSQDLFSVYLSADQD--SGSWVTFGGIDSSVYTGSLN	180
Db	189	LGFPIILSVEGVRPPLDVLVLEOGLDKPKPVSFYFNRPDPEVADGSELVLGGSDPAHYIPLT	248
QY	181	WVPVTVGYHQIIVDSITMNGEAIACAGCQAIVDTGTSLLTGPTSPIANIQSDIGASEN	240
Db	249	FVPVTPVAYQWIMHRVKVSRUTLCAQGCQAALIDTGTVPVIGPTEIRALHAAGIPL	308
QY	241	SGDMVMVSCSAISLPDIFTINGVQVPPPSAYILQ-SEGS--CISGFGMNNLPESG	296
Db	309	LAGEYIIRCSEIKPLPAVSLIIGVWFNLTAQDYVVIQFAGQDVRCLSGFRAUDIASPPV	368
QY	297	ELMWLGDFVIRQYFTTFVDR-----ANNQVGLA	323
Db	369	PVMILGDVFLGAYTVTFDRGDMKSGARVGLA	399

```

REPLY 15
US-08-974-691-2
; Sequence 2, Application US/08974691
; Patent No. 6225103
; GENERAL INFORMATION:
; APPLICANT: Keolsch, Gerald
; APPLICANT: Lin, Xinli
; APPLICANT: Tang, Jordan
; TITLE OF INVENTION: Cloning and Characterization of Napsin
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patricia L. Pabst
; STREET: 2800 One Atlanta Center, 1201 W. Peachtree
; STREET: St.
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentLin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/974,691
; FILING DATE: 20-NOV-1997
; CLASSIFICATION:

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/031,196
; FILING DATE: 20-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/046,126
; FILING DATE: 09-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: OMRF 166
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-873-8794
; TELEFAX: 404-873-8795
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 451 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
;
US-08-974-691-2

```

```

Query Match      39.1%; Score 668; DB 4; Length 451;
Best Local Similarity 41.1%; Pred. No. 1.3e-51;
Matches 136; Conservative 60; Mismatches 123; Indels 12; Gaps 5;

QY 5 PLENYLDMEYFTGIGTGAQFTVVFDGSSNLWVPVYCS--SLACTNHNRFNPEDSS 62
DB 69 PLSKFLDAQYFGEIGLGTTPQFTVAFDTGSSNLWVPVSRCHFFSVPCWFHFRFNPASS 128
QY 63 TVQSTSETVSYITGYSMTGILGYDTVOVGGISDTNQIFGLSETEPGSFLYYAPFDGIIG 122
DB 129 SFRPCKTFAIQYGRVDRVGLSEDKLTIGGKGVSVIFGEALWESSLVFTVSRPDGIIG 188
QY 123 LAYPSISSSGATPVFDNIWNOGLVSQDLFSVYLSADDQ--SGSVVIFGGIDSSYYTGSIN 180
DB 189 LCFPILSVGVRRPLDLVLEQGLDQKPVFSFYFNDRDPEVADGGELVGGSDPAHYIPPLT 248
QY 181 WVPVTEGYWQITVDSITWNGEAIACAECCQAIQVDTGTSLLTGPTSPIANIOSDIGASEN 240
DB 249 FVPVTPVAYWQHMERVKVGSRLTLCAQCCAAILDGTGTPVIVGPTTEIRALHAAIGGIPL 308
QY 241 SGGDMVYSCSAISSLPDVIPTFTINGVOYVPVPSAYTIQ--SEGS--CISGFGMNLPTESG 296
DB 309 LAGEYIIRCSEIPKLPVAVSLLIGGVWFNLTAQDYVIQFAQGDVRLCLSGFRALDIASPPV 368
QY 297 ELWILGDVFIQYFTVFDK---ANNQVGLA 323
DB 369 PWVILGDVFLGAYVIVFDRGDMKSGARVGLA 399

```

Search completed: September 6, 2001, 16:39:41
Job time: 130 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OW protein - protein search, using sw model

Run on: September 6, 2001, 16:43:32 ; Search time 225.25 Seconds
(without alignments)
87.740 Million cell updates/sec

Title: US-09-603-713-31

Perfect score: 1708

Sequence: 1 VDEQPLENYLDMEYFETIGI.....ROYFTVDFRANNOVGLAPVA 326

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 50623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_0601.*

```
1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1708	100.0	326	22	AA1986589 Human pepsin. Hom
2	1708	100.0	326	22	AA1986589 Pepsin protein. H
3	1074.5	62.9	388	20	AA1986589 Cat pregnancy asso
4	1010.5	59.2	365	10	AA1986589 Prochymosin. AAP
5	1008.5	59.0	390	22	AA1986589 Bovine chymosin po
6	1008.5	59.0	458	10	AA1986589 BamHI/SalI insert
7	1008.5	59.0	545	20	AA1986589 Oleosin-spacer-Met
8	1001.5	58.6	365	4	AA1986589 Sequence encoded b
9	1001.5	58.6	365	11	AA1986589 Sequence of calf p
10	1001.5	58.6	375	5	AA1986589 Sequence encoded b
11	1001.5	58.6	381	4	AA1986589 Sequence encoded b

12	999.5	58.5	381	4	AA1986589 Sequence encoded b
13	998.5	58.5	381	5	AA1986589 Sequence of a poly
14	996.5	58.3	381	5	AA1986589 Sequence of rennin
15	995.5	58.3	380	3	AA1986589 Pre-prorennin-A pr
16	991.5	58.1	379	4	AA1986589 Sequence encoded b
17	987.5	57.8	450	10	AA1986589 Sequence encoded b
18	967.5	56.6	381	13	AA1986589 Prochymosin (prore
19	911	53.3	375	20	AA1986589 Bovine pregnancy a
20	908.5	53.2	387	20	AA1986589 Bovine pregnancy a
21	880	51.5	391	20	AA1986589 Bovine pregnancy a
22	876	51.3	376	20	AA1986589 Bovine pregnancy a
23	875	51.2	381	20	AA1986589 Bovine pregnancy a
24	871.5	51.0	380	20	AA1986589 Bovine pregnancy a
25	868.5	50.8	380	20	AA1986589 Bovine pregnancy a
26	854	50.0	381	20	AA1986589 Bovine pregnancy a
27	849.5	49.7	380	20	AA1986589 Bovine pregnancy a
28	845.5	49.5	380	20	AA1986589 Bovine pregnancy a
29	841.5	49.3	380	20	AA1986589 Bovine pregnancy a
30	839.5	49.2	376	20	AA1986589 Bovine pregnancy a
31	835.5	48.9	380	20	AA1986589 Bovine pregnancy a
32	830	48.6	380	20	AA1986589 Bovine pregnancy a
33	827.5	48.4	377	20	AA1986589 Bovine pregnancy a
34	826.5	48.4	380	20	AA1986589 Bovine pregnancy a
35	825.5	48.3	380	20	AA1986589 Bovine pregnancy a
36	825	48.3	349	20	AA1986589 Human protease HPR
37	822.5	48.2	412	16	AA1986589 Death death associ
38	822.5	48.2	412	19	AA1986589 Death associated p
39	822.5	48.2	412	20	AA1986589 Human tumour-assoc
40	822.5	48.2	412	21	AA1986589 Amino acid sequenc
41	808.5	47.3	379	20	AA1986589 Bovine pregnancy a
42	806.5	47.2	392	20	AA1986589 Bovine pregnancy a
43	803.5	47.0	380	20	AA1986589 Bovine pregnancy a
44	803	47.0	379	20	AA1986589 Bovine pregnancy a
45	803	47.0	379	20	AA1986589 Bovine pregnancy a

ALIGNMENTS

RESULT 1
AAB66589
ID AAB66589 standard; Protein; 326 AA.
XX
AC AAB66589;
DT 12-APR-2001 (first entry)
DE Human pepsin.
XX
KW Human; pepsin; memapsin 2; nootropic; neuroprotective;
KW amyloid precursor protein; APP; memapsin 2 inhibitor;
KW Alzheimer's disease.
XX
OS Homo sapiens.
XX
PN WO2001006655-A2.
XX
PD 04-JAN-2001.
XX
PF 27-JUN-2000; 2000WO-US17742.
XX
PR 28-JUN-1999; 99US-0141363.
PR 30-NOV-1999; 99US-0168060.
PR 25-JAN-2000; 2000US-0177836.
PR 27-JAN-2000; 2000US-0178368.
PR 08-JUN-2000; 2000US-0210292.
XX
PA (OKLA-) OKLAHOMA MEDICAL RES FOUND.
PA (UNII) UNIV ILLINOIS FOUND.
XX
PI Tang JIN, Hong L, Ghosh AK;
XX WPI; 2001-137933/14.
DR

XX Novel memapsin 2 inhibitors which bind to active site of memapsin 2
 PT having 2 catalytic aspartic residues and substrate binding cleft, used
 PT to treat Alzheimer's disease by blocking amyloid precursor protein
 PT cleavage -
 XX
 XX Disclosure; Page 85-86; 86pp; English.
 XX
 XX The present sequence is given in a specification relating to an inhibitor
 CC of catalytically active memapsin 2. The inhibitor binds to the memapsin 2
 CC active site, which is defined by the presence of two catalytic aspartic
 CC residues and a substrate binding cleft. The inhibitor is useful for
 CC the treatment and diagnosis of Alzheimer's disease. It is useful in
 CC screens for individuals with a genetic predisposition to Alzheimer's
 CC disease. The inhibitor is useful as a reagent for specifically binding to
 CC memapsin 2 or memapsin 2 analogues and for aiding in memapsin 2
 CC isolation, purification and characterisation.
 XX
 XX Sequence 326 AA;

Query Match 100.0%; Score 1708; DB 22; Length 326;
 Best Local Similarity 100.0%; Pred. No. 3.1e-135;
 Matches 326; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDEQPLENYLDMVEYFGTIGTGAQDFTVVFDGSSNLWVPSVYCCSLACTNHNRPED 60
 DB 1 vdeqplenyldmvefgtigtgaqdfvfvdtgssnlwvpsvycslactnhrfped 60
 QY 61 SSTQSTSETSYITGYGSMTGILGYDTVQVGGISDNTQIFGLSETPGSFLYYAPFDGI 120
 DB 61 sstqstsetsvitygtgsmtgilgydvtvqvggisdnqifglsetpgsfllyyapfdgi 120
 QY 121 LGLAYPSISSSGATPVFDNIWNGGLVSDQSFVLSADDSGSGSVVIFGGIDSSYYTGSIN 180
 DB 121 lglaypsissgatpvdnlnwngglvsgdlsfvlsaddsgsgsvvifggidssyytgsin 180
 QY 181 WVPVTVEGYWQITVDSITMNGEAIACAEQCAIVDTGTSLLTGTPTSPANTQSDIGASEN 240
 DB 181 wvpvtvegywqitvdsitmngaeaiacaeqcaivdtgtslltgtptspiani qsdigasen 240
 QY 241 SDGDMVVSCTSAISSLPDIFTINGVQYVPVPPSAVILQSEGSCISGFGGMNLPTESGELWI 300
 DB 241 sdgdmvvsctsaisslpdiftingvqvpvppsavilqsegscisgfggmnlptesgelwi 300
 QY 301 LGDVFIROYFTVFDNRANNQVGLAPVA 326
 DB 301 lgdvfirqyftvfdnranngvqlapva 326

RESULT 2
 AAB61351
 ID AAB61351 standard; protein; 326 AA.

XX AC AAB61351;
 XX
 DT 02-APR-2001 (first entry)
 XX Pepsin protein.
 DE
 XX Memapsin 2; catalyst; Alzheimer's.
 KW
 XX Homo sapiens.
 OS
 XX WO200100663-A2.
 PN
 XX 04-JAN-2001.
 PD
 XX 27-JUN-2000; 2000WO-US17661.
 PF
 XX 28-JUN-1999; 99US-0141363.
 PR 30-NOV-1999; 99US-0168060.
 PR 25-JAN-2000; 2000US-0177836.

PR 27-JAN-2000; 2000US-0178368.
 PR 08-JUN-2000; 2000US-0210292.
 XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.
 PA
 XX Tang JUN, Lin X, Koelsch G;
 PI WPI; 2001-102885/11.
 XX
 XX Purified recombinant catalytically active memapsin 2, used to screen
 PT inhibitors of it, which are used to treat and prevent Alzheimer's
 PT disease -
 XX
 XX Disclosure; Page 86-88; 86pp; English.
 XX
 XX The present invention relates to a purified recombinant
 CC catalytically active memapsin 2. The invention may be used for
 CC isolating inhibitors which are used to treat or prevent
 CC Alzheimer's disease. The invention may also be used to screen
 CC for individuals more genetically prone to develop Alzheimer's
 CC disease.
 XX
 XX Sequence 326 AA;

Query Match 100.0%; Score 1708; DB 22; Length 326;
 Best Local Similarity 100.0%; Pred. No. 3.1e-135;
 Matches 326; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDSQPLENYLDMVEYFGTIGTGAQDFTVVFDGSSNLWVPSVYCCSLACTNHNRPED 60
 DB 1 vdeqplenyldmvefgtigtgaqdfvfvdtgssnlwvpsvycslactnhrfped 60
 QY 61 SSTQSTSETSYITGYGSMTGILGYDTVQVGGISDNTQIFGLSETPGSFLYYAPFDGI 120
 DB 61 sstqstsetsvitygtgsmtgilgydvtvqvggisdnqifglsetpgsfllyyapfdgi 120
 QY 121 LGLAYPSISSSGATPVFDNIWNGGLVSDQSFVLSADDSGSGSVVIFGGIDSSYYTGSIN 180
 DB 121 lglaypsissgatpvdnlnwngglvsgdlsfvlsaddsgsgsvvifggidssyytgsin 180
 QY 181 WVPVTVEGYWQITVDSITMNGEAIACAEQCAIVDTGTSLLTGTPTSPANTQSDIGASEN 240
 DB 181 wvpvtvegywqitvdsitmngaeaiacaeqcaivdtgtslltgtptspiani qsdigasen 240
 QY 241 SDGDMVVSCTSAISSLPDIFTINGVQYVPVPPSAVILQSEGSCISGFGGMNLPTESGELWI 300
 DB 241 sdgdmvvsctsaisslpdiftingvqvpvppsavilqsegscisgfggmnlptesgelwi 300
 QY 301 LGDVFIROYFTVFDNRANNQVGLAPVA 326
 DB 301 lgdvfirqyftvfdnranngvqlapva 326

RESULT 3
 AAY32058
 ID AAY32058 standard; Protein; 388 AA.
 XX AC AAY32058;
 XX
 DT 05-JAN-2000 (first entry)
 XX Cat pregnancy associated glycoprotein (PAG).
 DE
 XX PAG; pregnancy associated glycoprotein; cat; diagnosis.
 KW
 XX Felis domestica.
 OS
 XX WO9947934-A2.
 PN
 XX 23-SEP-1999.
 PD
 XX 19-MAR-1999; 99WO-US06038.
 PF

QY 305 FIRQYFTVFDNRANNOVGLA 323
Db 437 fireyysvfdnrannlvglia 455

RESULT 7

RAY33830
ID AAY33830 standard; Protein: 545 AA.

XX AC AAY33830;

XX DT 29-NOV-1999 (first entry)

XX Oleosin-spacer-Met-prochymosin amino acid sequence.

XX oil-body; lipid body; oleosome; spherosome; separation;
KW fusion protein; heterologous polypeptide; commercial production;
KW plasmid; oil-body targeting sequence.

XX Synthetic.

OS Arabidopsis thaliana.

XX Key Location/Qualifiers

FT Protein 1..118
FT /label= Oleosin targeting sequence

FT 119..175

FT /label= Oleosin protein

FT cleavage_site 176..181

FT /note= "Thrombin cleavage site"

FT Protein 182..483

FT /label= Met-Prochymosin

XX US5948682-A.

PN 07-SEP-1999.

XX 25-APR-1997; 97US-0846021.

XX 25-APR-1997; 97US-0846021.

XX 22-FEB-1991; 91US-0659835.

XX 16-NOV-1993; 93US-0142418.

XX 30-DEC-1994; 94US-0366783.

XX (SEMB-) SEMBIOSYS GENETICS INC.

PA Moloney MM;

PI WPI; 1999-517960/43.

XX N-PSDB; AA208463.

XX Expression of a heterologous polypeptide on an oil body protein is
XX useful for the production of e.g. enzymes, antibodies, hormones

XX Example 10; Fig 6; 48pp; English.

XX This is the amino acid sequence of a HindIII fragment containing the
CC oleosin-spacer-Met-prochymosin sequence. This HindIII fragment was
CC joined to a nopaline synthase terminator and cloned into binary vector
CC pCGN1559. The resulting plasmid was called pBSORTPT and introduced
CC into A.tumefaciens. The resulting bacterial strain was used to transform
CC B.napus plants.
CC The DNA which encodes a chimeric fusion protein that consists of the
CC oil-body targeting sequence, a transcription regulation sequence and the
CC DNA of the protein of interest can be used to produce antibodies,
CC glycanases, hormones, proteases, protease inhibitors, seed storage
CC proteins, thrombin inhibitors, hirudin, interleukins, chymosin, cystatin,
CC xylanase, carp growth hormone, zein or a collagenase.
CC The enzyme may be cleaved from the oil body protein or used in
CC association with the oil body fraction.
CC Allows production of commercially important proteins on a superior scale
CC to production by conventional systems. The expressed heterologous
CC protein can be easily separated from host cell components
CC due to the use of the oil body as a carrier protein.

XX SQ Sequence 545 AA;

Query Match 59.0%; Score 1008.5; DB 20; Length 545;

Best Local Similarity 59.6%; Pred. No. 2.4e-76;

Matches 190; Conservative 43; Mismatches 81; Indels 5; Gaps 2;

QY 5 PLENYLDMEYFGTIGTIGTTPAODFTVVDFTGSSNLWVPVSVYVYSSSLACTNHNHNPEDSSSTY 64

Db 229 pltnyldsqyfgkylgtppqeftvldtgssdfwvpslycksnacknhqridpkrkstf 288

QY 65 QSTSETVITYGTGSMTGILGYDVTVOVGISDTNIFGLSETEPGSFYLYAFPDPGILGLA 124

Db 289 qnlgkplshygtgsmqgilgydvtvsnivdiqvtglstdepgdvfyaeafdgilgma 348

QY 125 YPSISSSGATPVFDNIWNOGLVSDQLFSVYLSADDOGSVWIFGIDSSYYTGSNLWVPV 184

Db 349 ypslaseysipvfdnmnrhlvaqdlfsvymdrngge-smiltgaldpssytcshwvvpv 407

QY 185 TVEGYWQITVDSITMNGEAIACAEQCQAIIVDTGTSLLTGPTSPIANIOSDIGASENSDGD 244

Db 408 tvqyqwqftvdsvtisgvvaceggcqaiddtqtsklvgpssdlinlqalgatqnqyge 467

QY 245 MVVSCSAISSLPDIIVFTINGVQVPVPPSAYILQSESGISGFGQNMNLPTEGELWILGDV 304

Db 468 fdidcdnlisymptvveingkmypitpsaytsqdgqfctsgfqsan----hsqkwilgdv 523

QY 305 FIRQYFTVFDNRANNOVGLA 323

Db 524 fireyysvfdnrannlvglia 542

RESULT 8

RAP30603

ID AAP30603 standard; Protein: 365 AA.

XX AC AAP30603;

XX DT 14-JUN-1992 (first entry)

XX Sequence encoded by prorennin cDNA in pCR 10001.

XX Rennin; renln; enzyme; protease.

XX Bos taurus.

XX EP73029-A.

XX 02-MAR-1983.

XX 19-AUG-1982; 82EP-0107601.

XX 24-AUG-1981; 81JP-0131631.

XX (BEPP/) BEPPU T.

XX Beppu T, Uozumi T, Nishimori K;

XX WPI; 1983-22976K/10.

XX N-PSDB; AAN30063.

XX Plasmid contg. calf pro:rennin DNA - and transformed

XX microorganisms

XX Example; Page 20-23; 32pp; English.

XX The inventors claim recombinant plasmids contg. the cDNA of calf

XX prorennin. Specified plasmids are pCR 10001 and pCR2001 (contg. the

XX whole sequence plus the lac promoter region). Also new are

XX microorganisms transformed with the plasmids, esp. E. coli CRI (ATCC

XX 391710) contg. plasmid pCR2001.

XX

5	PLENYLDMEYFGTIGTTPAQDFTVVFDTGSSNLWVPSVYCSLACTNHNRENPEDSSTY	64	5	PLENYLDMEYFGTIGTTPAQDFTVVFDTGSSNLWVPSVYCSLACTNHNRENPEDSSTY	64
49	pltnyldsqyfgkiylgtppqeftvlfdtgssdfwvpsiycksnacknhqfdrprkstf	108	49	pltnyldsqyfgkiylgtppqeftvlfdtgssdfwvpsiycksnacknhqfdrprkstf	108
65	QSTSEVTSITVTGSGMTGILGYDVTVOVGISDTNQIFGLSETEPGSFLYAPDGLGLA	124	65	QSTSEVTSITVTGSGMTGILGYDVTVOVGISDTNQIFGLSETEPGSFLYAPDGLGLA	124
109	qnlqkplshygtgsmqgilgydvtvsnlvdigtvlgstqpgdvftyaefdgilgma	168	109	qnlqkplshygtgsmqgilgydvtvsnlvdigtvlgstqpgdvftyaefdgilgma	168
125	YPSISSSGATVPFDINWQGLSVLSDQSGSVVIFGGIDSSYYTGSLNWVPV	184	125	YPSISSSGATVPFDINWQGLSVLSDQSGSVVIFGGIDSSYYTGSLNWVPV	184
169	yplaseyslpvfdmnmrhvagdflsvymdrngqe-smiltgaldpsyytgsllhwvpv	227	169	yplaseyslpvfdmnmrhvagdflsvymdrngqe-smiltgaldpsyytgsllhwvpv	227
185	TVEGYWQITVDSITMNGEAIACACQAIIVDTGTSLLTGTPSIANIQSDIGASENSDGD	244	185	TVEGYWQITVDSITMNGEAIACACQAIIVDTGTSLLTGTPSIANIQSDIGASENSDGD	244
228	tvqgywqftvdsitgsvvacegcqaildgtsklvpgssdlnlqgaigatqngyde	287	228	tvqgywqftvdsitgsvvacegcqaildgtsklvpgssdlnlqgaigatqngyde	287
245	MVYSCSAISLPIVFTINGVQVPPPSAYILQSEGSCISGFGQMNLPTESGELWILGDV	304	245	MVYSCSAISLPIVFTINGVQVPPPSAYILQSEGSCISGFGQMNLPTESGELWILGDV	304
288	fdidcdnlsympvtvfeingkmypitpsaytsdqgfgctsgfqsene---hsqkwilgdv	343	288	fdidcdnlsympvtvfeingkmypitpsaytsdqgfgctsgfqsene---hsqkwilgdv	343
305	FIROYFTVFDNRANNOVGLA 323		305	FIROYFTVFDNRANNOVGLA 323	
344	fireyysvfdannlvglgla 362		344	fireyysvfdannlvglgla 362	
RESULT 10			RESULT 9		
AAP40078			AAR05080		
ID AAP40078 standard; Protein; 375 AA.			ID AAR05080 standard; protein; 365 AA.		
XX AC AAP40078;			XX AC AAR05080;		
XX DT 02-FEB-1992 (first entry)			XX DT 05-OCT-1990 (first entry)		
XX DE Sequence encoded by prochymosin gene.			XX DE Sequence of calf pro-rennin.		
XX KW Prochymosin expression vector; E.coli trp operon; chymosin.			XX KW Pro-rennin; ds.		
XX OS Bos taurus.			XX OS Bos taurus.		
XX PN EPI121775-A.			XX PN JP02109984-A.		
XX PD 17-OCT-1984.			XX PD 23-APR-1990.		
XX PF 07-MAR-1984; 84EP-0102451.			XX PF 01-JAN-1988; 88JP-0302176.		
XX PR 09-MAR-1983; 83JP-0038439.			XX PR 01-JAN-1988; 88JP-0302176.		
XX PA (BEPP) BEPPU T.			XX PA (VEPP) VEPPU T.		
XX PI Beppu T, Uozumi T, Nishimori K, Shimizu N, Kawaguchi Y, Hidaka M;			XX PI WPI; 1990-168358/22.		
XX PS WPI; 1984-258001/42. N-PSDB; AAN40055.			XX PS N-PSDB; AAQ04683.		
XX PT Expression plasmid comprising prochymosin gene and vector - useful for transforming Escherichia coli for prochymosin prodn.			XX PT Complex plasmid and microbe - contains calf pro-rennin cDNA.		
XX PS Disclosure; Fig 1; 59pp; English.			XX PS Disclosure; 32; 13pp; Japanese.		
XX CC The inventors claim the prochymosin gene comprising a nucleotide sequence from (a) the 1st codon (GCT) to the 365th codon (ATC), or (b) the 5th codon (CGG) to the 365th codon (ATC); and recombinant plasmids harboured by Escherichia coli strains deposited as FERM BP-262, -263 and -264. Any portion of the nucleotide sequence as described in AAN40055 can be used. Also claimed is a vector derived from plasmid pBR322. Typically plasmid pCR501 is obtd. from pOCT 2. The transcriptional direction of pOCT 3 is opposite to that of pOCT 2; it is clockwise in pOCT 2 (5' to 3') whereas counter clockwise in pOCT 3.			XX CC Protein product may be expressed in E.coli expression system from plasmid pBR322.		
XX CC Sequence 375 AA;			XX CC Sequence 365 AA;		
XX SQ			XX SQ		
Query Match			Query Match		
Best Local Similarity 58.6%; Score 1001.5; DB 4; Length 365;			Best Local Similarity 59.2%; Score 1001.5; DB 11; Length 365;		
Matches 189; Conservative 43; Mismatches 82; Indels 5; Gaps 2;			Matches 189; Conservative 43; Mismatches 82; Indels 5; Gaps 2;		

Query Match	58.6%	Score	1001.5;	DB 5;	Length	375;			
Best Local Similarity	59.2%;	Pred. No.	5.6e-76;						
Matches	189;	Conservative	43;	Mismatches	82;	Indels	5;	Gaps	2;
<hr/>									
QY	5	PLENYLDMEVFGTIGTGAQDFTVVFDTGSSNLWVSVYCSSLACINHNHNFNPEDSSTY	64						
Db	59	pltnyldsgfkglyigtbpqetvldtssdfwpslycksnacknhgrfdprkastf	118						
QY	65	QSTSETSVINYGSGMFGILGYDTPVQGGI:SDTNQIFGLSETPEPSFLYVAPFGSLGLA	124						
Db	119	qnlgkplshlygtsgmgilgydvtvsnvldiqtvglstqebgdvftyaefdgilgma	178						
QY	125	YPTSSSGANPVDNINWQGLVSQDLFSVYLSADQSGSVVIFGIDSSVYTGSLNWVPV	184						
Db	179	ypslaseyslpvfannmnrhlhvaqdltsvymdrngqe-smltgalgopsyvtgshwvpv	237						
QY	185	TVBEGYMQITVDSITMNGEACIACAECCQAIVDTGTSLLTGTSPIANQSDICASENSDGD	244						
Db	238	tvqgyqgftvdsvtsgvvvacggcqaalidtgtskivgssdlhnlqalgaatcqnqyde	297						
QY	245	MVYSCSAISSLPDIVFTINGVQVPVPPSPYILASEGCSIGSQFGQGNLPTEGSELWILGDV	304						
Db	298	fdidcdnlstympcvvfeingkmkmyltpsaytsqqgfcctagfqsen----hsqkwilgdv	353						
QY	305	FIHQYTFVFDNRANQVCLA	323						
Db	354	fireyysvfdannlvqla	372						

RESULT 11
AAP30086
ID AAP30086 standard; Protein; 381 AA.
XX
XX AAP30086;
AC
XX
XX
14-JUN-1992 (first entry)
DT
DT
XX
XX
DE Sequence encoded by cDNA sequence corresponding to one of the
DE allelic forms (B) of bovine preprochymosin.
DE
XX
XX Chymosin; enzyme; rennet; cheese.
KW

XX	Bos taurus.		
XX			
XX	Key	Location/Qualifiers	
XX	Peptide	1..16	
FT	FT	/label= signal	
XX			
XX	EP77109-A.		
PN			
XX			
XX	20-APR-1983.		
PD			
XX			
XX	13-OCT-1982;	82EP-0201272.	
PF			
XX			
XX	14-OCT-1981;	81GB-0031004.	
PR			
XX			
XX	(UNIL) UNILEVER NV.		
PA			
XX			
PI	Maat J, Verrips CT, Ledeboer AM, Edens L;		
XX			
XX	WPI; 1983-39656K/17.		
DR			
XX	N-PSDB; AAN30049.		
DR			
XX			
XX			
PT	DNA molecules comprising genes for preprochymosin - used to		
PT	transform microorganisms to give strain producing the		
PT	prepro-enzyme and its allelic and maturation forms		
XX			
XX			
PS	Claim 2; Fig 1; 53pp; English.		

Preprochymosin is an intermediate (via prochymosin and pseudochymosin) for the enzyme chymosin, which is the essential milk-clotting component of rennet and is used in cheese manufacture.

CC	AAN30049 corresp. to mRNA isolated from the fourth stomach of a	
CC	peruminant calf (abomasum, Frisian cow).	
XX		
SQ	Sequence	381 AA;
	Query Match	58.6%; Score 1001.5; DB 4; Length 381;
	Best Local Similarity	59.2%; Pred. No. 5.7e-76;
	Matches 189; Conservative 42; Mismatches 83; Indels 5; Gaps:	
QY	5	PLENYLDNEYFGIGIGTTPAQDFTVVFDTGSSNLWSPVSCSLACTNHNRFNPEDSSTY 64
DB	65	pltnldsqyfgklylgtppqetfvidtgssdfwpslycksnacknhrfdpkrkstf 124
QY	65	QSTSETVSYNYGTGSGTWGLIGYDTVQVGGISDNTNQIFGLSETPEGSFLYAPFDGILGLA 124
DB	125	qnlgkplslhygtgsmgqilgydtvtvsnlvdilqgtvgilstqpgdvfcyaefdgilgma 184
QY	125	YPSITSSGAPVPFNDTINWQGLFVSQDLFVSVLSADDSQGSVVIFGGIDSSYYTGSLSNWVPV 184
DB	185	ypslasqyslpvfdnmnmshlvaqdlfsvymdrnge-smiltgaldpssyqslhwvpy 243
QY	185	TVGQYWQITVDSITMNGEAIACAEQCOAIVDTGTSLLTGTPSPIANIQSDIGASENSDGD 244
DB	244	tvqgywqftvdsvtisgvvaceggcqailldtgsklvgspsdillnlgqalgtatqnyge 303
QY	245	MVYSCSAISSLPDITVFITTINGVQVPPVPSAYILQSEGCSTGGFGMKLPETSECELWITLGV 304
DB	304	fdldcdnlsmptvfvfeingkmvpltpsaytsqdqgfcfctsaqfen----haqkwilgav 359
QY	305	FIRQYFTVFDRAANNQVGLA 323
DB	360	fireyysvfdraannlvqla 378

RESULT	12
AAP30446	
ID	AAP30446 standard; Protein; 381 AA.
XX	
AC	AAP30446;
XX	
DT	03-AUG-1992 (first entry)
XX	
DE	Sequence encoded by preprothymosin cDNA.
XX	
KW	Milk-clotting; cheese making; enzyme; zym
XX	
OS	Cow.

Key	Location/Qualifiers
Region	1..16
FT	/label= preprochymosin
FT	17..58
FT	/label= prochymosin
FT	59..381
FT	/label= chymosin

AA	
PN	GB2100737-A.
XX	
XX	06-JAN-1983.
PD	
XX	
XX	11-JUN-1982; 82GB-0017096.
PF	
XX	
PR	10-FEB-1982; 82GB-0003907.
XX	
PA	(CELL-) CELLTech LTD.

XX	Carey NH, Harris TJR, Lowe PA, Doel MT, Entage JS;
PI	
XX	
XX	WPI: 1983-00545K/01.
DR	N-PSDB; AAN30209.
XX	
XX	Prodn. or calf stomach chymosin for cheese making - by
PT	

PT cultivation of micro-organisms transformed with vector system
XX
XX
PS Claim 41; Fig 4; 26pp; English.
XX
XX
CC The inventors claim a method for the prodn. of calf stomach chymosin
CC for cheese making. Genes and polypeptides for preprochymosin,
CC prochymosin and chymosin are claimed, as are vector systems and a
CC prochymosin primer.
XX
XX Sequence 381 AA;
SQ

Query Match	58.5%	Score	999.5;	DB	4;	Length	381;
Best Local Similarity	59.2%;	Pred.	No. 8.4e-76;				
Matches	189;	Conservative	43;	Mismatches	82;	Indels	5; Gaps
QY	5	PLENLDMEYFGTIGITPAQDFTVFDPTGSSNLWVPWSVCSSLACTNHNRFNPEDSSTY	64				
Db		: : : : : : : : : : : :					
QY	65	pltnldsqyfgkylgtpbpqetvifdgtssdfwpslycksnacknhqfdrprkstf	120				
Db		: : : : : : : : : : : :					
QY	65	QSTSETVTYTGTSGMTGLGYDTVOVGGISDPNQIFGLSETEPGSFLYYAPFDCILGLA	120				
Db		: : : : : : : : : : : : :					
QY	125	qnlgkplsibhytgsmagilgydtvtvsnlvdltqvcvgistgdvfyaeafdgllma	180				
Db		: : : : : : : : : : : : : :					
QY	125	YPSTSSGATPVFNINWQGLSVQDLFSVYLSDAQDSGSVFVFGIDSYTYTGSLNWVPV	180				
Db		: : : : : : : : : : : :					
QY	185	yplslaseysipvfvdmmnrhlvaqdlfsvyymdrigge-smitgainpsyytgslnhwvp	243				
Db		: : : : : : : : : : : :					
QY	185	TVEGYWLTVDSTIMNGEATACAEQGCOAIVDRGTSLLTGPTSPITANGSDIGASENSDG	244				
Db		: : : : : : : : : : : :					
QY	244	tvqygwfctvdsvlisgvvvaceggccalldtgtkslvgpssdlinlqgalgaicnqayge	303				
Db		: : : : : : : : : : : :					
QY	245	MVWSCSAISSLPDIPTFTINGVQVPVPVSPILOESGCISGFCQMNPETSEGELWLIGDV	304				
Db		: : : : : : : : : : : : : :					
QY	304	fddcdnlsympvtvfeingkmpltpsaytesqdggfcftsgfsen---hsqkwilgdv	359				
Db		: : : : : : : : : : : :					
QY	305	FIRQYFTVFDRAANNQVLA	323				
Db		: : : : : : : : : : : :					
QY	360	fireysvfdraannlvla	378				
Db		: : : : : : : : : : : :					

RESULT	13
AAP40559	
ID	AAP40559 standard; Protein; 381 AA.
XX	
AC	AAP40559;
XX	
TT	04-FEB-1992 (first entry)
XX	
EE	Sequence of a polypeptide displaying milk clotting activity.
XX	
WW	Cheese-making; recombinant protein; rennet substitute; milk clot.
XX	
SS	Bos taurus.
XX	
H	Key
T	Peptide
T	Location/Qualifiers
T	1..16
T	/label= signal
T	Protein
X	17..381
N	
N	EP123928-A.
X	
D	07-NOV-1984.
X	
F	30-MAR-1984; 84EP-0103551.
X	
R	31-MAR-1983; 83US-0480860.
R	
X	(CODO-) CODON GENETIC ENG.
A	
A	
I	Cashion LM, McCaman MT, Rice CW, Sias SR;
X	WPI; 1984-277277/45.

DR	N-PSDB; AAN40295.
XX	
PT	Recombinant DNA coding for milk clotting polypeptide - which is
PT	expressed in transformed bacteria
XX	
PS	CladM 10; Fig 2; 39pp; English.
XX	
CC	Also claimed is E.coli JM83/plc7 (ATCC 39325) which is transformed
CC	with plc7 contg. the prorennin derived sequence fused in phase with
CC	B-galactosidase. The plc7 prorennin expression plasmid includes
CC	sequences which code for both the pseudorennin and mature rennin
CC	cleavage sites between AAs 28-29 and AAs 42-43, respectively.
XX	
SQ	Sequence 381 AA:

Query Match	58.5%	Score 998.5	DB 5	Length 381
Best Local Similarity	58.9%	Pred. No. 1e-75		
Matches 188	Conservative 44	Mismatches 82	Indels 5	Gaps
QY	5	PLENYLDMEYFETIGCTPAQDFVVFVFGSSNLWPSVYCSSLACTNHNHNRFPEDSSY	64	
Db	65	pltnyldsqyfgkiylgtppqeftvldtgsdfwpslycksnacknhqfdrkfstf	124	
QY	65	QSTSTSVTSTYGTGSMTCGILGYDVTQVQGGISDTHQIFGLSETEPGSFLEYAPFDGILGLA	124	
Db	125	qnlgkplsihygtsgmgilgydvtvsnvldiqetvglstqpgdvyftyaefdgllilma	184	
QY	125	YPTSSSSGATPVFDNIWNQGLVSODLFSVYLASDDQSGSVVIFGIDSSYTTGSLNWVPV	184	
Db	185	ypslaseysipvfdnmnrhivagdlfsydnrngqe-smiltiga dpsyvtgslhwvpy	243	
QY	185	TVEGYWQITVDSIITWNGEATACAGCOAIVDTGTSLLTGPTSPITANIQSDGASENSDGD	244	
Db	244	tvqgywqfvdsvtsgvwwvaceggccalltdgtkslvgpssdillnqaiqatqnyde	303	
QY	245	MVWSCSAISSLPDILVETINGVQVPPPSAVILQSEGSCISGFGMNLPTESGBELWILGDV	304	
Db	304	fdidcdnlsmptvvtfeingkmypiltpsaytsdqdgfctsgfgsen----hsqkwilgdv	359	
QY	305	FIROYTFVDRANNQVGLA	323	
Db	360	firoysvfdrrannlvqla	378	

RESULT 14	
AAP40218	
ID	AAP40218 standard; Protein; 381 AA.
XX	
AC	AAP40218;
XX	
DT	25-JAN-1992 (first entry)
XX	
DE	Sequence of rennin encoded by recombinant CGF4.
XX	
KW	Yeast expression vector; GALL promoter; <i>Saccharomyces cerevisiae</i> .
XX	
OS	<i>Bos taurus</i> .
XX	
PN	GB2137208-A.
XX	
PD	03-OCT-1984.
XX	
PF	28-FEB-1984; 84GB-0405129.
XX	
PR	28-FEB-1983; 83US-0470911.
XX	
PA	(COLB) COLLABORATIVE RES INC.
XX	
PI	Botstein D, Davis RW, Fink GR, Taunton-Rigby A, Knowlton RG;
PPI	Mao JI, Moir DT, Goff CG;
XX	
DR	WPI; 1984-245517/40.

DR N-PSDB; AAN40180.
 XX
 PT DNA segment contg. GAL1 promoter linked to gene - useful for
 PT direction of expression of the gene in yeast cell
 XX
 XX Example; Table 4, Page 21-23; 35pp; English.
 PS
 XX The inventors claim a DNA segment contg. GAL1 promoter linked to
 CC gene - useful for direction of expression of the gene in yeast cell.
 CC The recombinant material carrying a GAL1 promoter of the yeast
 CC galactokinase gene may be used in expressing a desired protein, esp.
 CC bovine growth hormone, interferon, prorennin or prorennin, in the
 CC yeast cell. Strains of Saccharomyces cerevisiae producing the
 CC polypeptides are produced. Yeast strains deposited as ATCC 20643,
 CC 20661, 20662 and 20663, strain designations CGY 196, 457, 461 and
 CC 528, resp. are new.
 XX
 XX Sequence 381 AA;

Query Match 58.3%; Score 996.5; DB 5; Length 381;
 Best Local Similarity 59.2%; Pred. No. 1.5e-75;
 Matches 189; Conservative 42; Mismatches 83; Indels 5; Gaps 2;
 QY 5 PLENYLDMEYFGTIGTGAQDFTVVFDGSSNLWVPSVYCSLACTNHNRFNPEDSSTY 64
 DB 65 pltnyldsqyfgkiyigtppqefvifdtgssdfwvpslycksnacknhqrfdrkstf 124
 QY 65 QSTSEIVSYTYGTGSMGTILGVDTVQVGGISDTNQIFGLSETEPGSFLYAPDGLGLA 124
 DB 125 qnlgkplshygtgsmqgilgdytvtvsnivdiqctvgistqepgdvfyaeafdgllgma 184
 QY 125 YPSISSSGATPVFDNIWNOGLVSQDLFSVYLSADDSGVSIVFGGIDSSVYTGSLNWVPV 184
 DB 185 ypsraseysipvfdmmnrhlvaqdlfsvymdrnge-smiltlgaidpsyytgsllhwvpv 243
 QY 185 TVEGYWQITVDSITMNGEAIACAEQCAIVDGTSLTGTPTSPIANIQSDIGASENSDGD 244
 DB 244 tvqgywqfvdsvtisgvvvaceggcqailldtgtsklvgpsdillniqgaigatqnqyde 303
 QY 245 MVVSCSAISLPDIPVTVINGVOYVPPPSAYILQSEGSCISGFGMNLPTESGELWILGDV 304
 DB 304 fdlcdnlisymptvfeingmkypltpsaytsqdgqgfcstsgfqsense- ---hsqkwilgdv 359
 QY 305 FIRQYETVFDNRANNOVGLA 323
 DB 360 fireyysvfdannlvla 378

RESULT 15
 AAP20038
 ID AAP20038 standard; Protein; 380 AA.
 XX
 AC AAP20038;
 XX
 DT 16-DEC-1992 (first entry)
 XX
 DE Pre-prorennin-A protein sequence.
 XX
 KW Pre-pro-rennin; rennin; prorennin; enzyme; EC-3.4.23.4; chymosin;
 KW protease; milk-clotting enzyme; ss.
 XX
 OS Bos taurus.
 XX
 PN GB2091271-A.
 XX
 PD 28-JUL-1982.
 XX
 PF 15-JAN-1982; 82GB-0001120.
 XX
 PR 01-DEC-1981; 81US-0325481.
 PR 16-JAN-1981; 81US-0225717.
 XX

PA (COLB) COLLABORATIVE RES INC.
 XX
 PI Alford BL, Mao J, Moir DT;
 XX
 DR WPI: 1982-62028E/30 (62028E).
 DR P-PSDB; AAP20038.
 XX
 PT Transformed cells producing rennin and its precursors - contg.
 PT appropriate recombinant DNA material
 XX
 PS Disclosure; Table 1; 39pp; English.
 XX
 CC DNA sequences either side of the protein sequence
 CC can be removed and are not essential to use of the gene in
 CC expression. The protein may be expressed in E. coli using
 CC plasmid pCE21. The resulting expressed enzyme is a well
 CC known milk-clotting enzyme used in cheese-making.
 XX
 XX Sequence 380 AA;

Query Match 58.3%; Score 995.5; DB 3; Length 380;
 Best Local Similarity 58.9%; Pred. No. 1.8e-75;
 Matches 188; Conservative 43; Mismatches 83; Indels 5; Gaps 2;
 QY 5 PLENYLDMEYFGTIGTGAQDFTVVFDGSSNLWVPSVYCSLACTNHNRFNPEDSSTY 64
 DB 64 pltnyldsqyfgkiyigtppqefvifdtgssdfwvpslycksnacknhqrfdrkstf 123
 QY 65 QSTSEIVSYTYGTGSMGTILGVDTVQVGGISDTNQIFGLSETEPGSFLYAPDGLGLA 124
 DB 124 qnlgkplshygtgsmqgilgdytvtvsnivdiqctvgistqepgdvfyaeafdgllgma 183
 QY 125 YPSISSSGATPVFDNIWNOGLVSQDLFSVYLSADDSGVSIVFGGIDSSVYTGSLNWVPV 184
 DB 184 ypsraseysipvfdmmnrhlvaqdlfsvymdrnge-smiltlgaidpsyytgsllhwvpv 242
 QY 185 TVEGYWQITVDSITMNGEAIACAEQCAIVDGTSLTGTPTSPIANIQSDIGASENSDGD 244
 DB 243 tvqgywqfvdsvtisgvvvaceggcqailldtgtsklvgpsdillniqgaigatqnqyde 302
 QY 245 MVVSCSAISLPDIPVTVINGVOYVPPPSAYILQSEGSCISGFGMNLPTESGELWILGDV 304
 DB 303 fdlcdnlisymptvfeingmkypltpsaytsqdgqgfcstsgfqsense- ---hsqkwilgdv 358
 QY 305 FIRQYETVFDNRANNOVGLA 323
 DB 359 fireyysvfdannlvla 377

Search completed: September 6, 2001, 16:43:33
 Job time: 362 sec

us-09-603-713-31.rag

Fri Sep 7 10:58:28 2001

Fri

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:49:57 ; Search time 231.42 Seconds
(without alignments)
5.717 Million cell updates/sec

Title: US-09-603-713-30
Perfect score: 46
Sequence: 1 VSGVLLSRK 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SPTREMBL_16:*
- 2: sp-archaea:*
- 3: sp-bacteria:*
- 4: sp-fungi:*
- 5: sp-human:*
- 6: sp-invertebrate:*
- 7: sp-mammal:*
- 8: sp-mhc:*
- 9: sp-organelle:*
- 10: sp-phage:*
- 11: sp-plant:*
- 12: sp-rodent:*
- 13: sp-unclassified:*
- 14: sp-vertebrate:*
- 15: sp-virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	38	82.6	487	2 Q9L3L1	Q9L3L1 salmonella
2	36	78.3	709	5 Q9VK84	Q9VK84 drosophila
3	36	78.3	1194	13 Q9W737	Q9W737 gallus gall
4	35	76.1	670	10 Q9FF80	Q9FF80 arabidopsis
5	35	76.1	1505	2 Q9F0D7	Q9F0D7 streptomyce
6	35	76.1	4247	2 Q9L8H4	Q9L8H4 streptomyce
7	34	73.9	80	5 Q23577	Q23577 caenorhabdi
8	34	73.9	204	1 Q9HSG4	Q9HSG4 halobacteri
9	34	73.9	366	1 Q9V124	Q9V124 pyrococcus
10	34	73.9	528	5 Q9UAF9	Q9UAF9 ephydatia f
11	33	71.7	82	2 Q9X0T3	Q9X0T3 thermotoga
12	33	71.7	182	13 Q9W663	Q9W663 trachemys s
13	33	71.7	182	13 Q9W661	Q9W661 trachemys s
14	33	71.7	182	13 Q9W660	Q9W660 poephila gu
15	33	71.7	182	13 Q9W659	Q9W659 poephila gu
16	33	71.7	182	13 Q9W657	Q9W657 coryphaenoi
17	33	71.7	182	13 Q9W655	Q9W655 carassius a
18	33	71.7	182	13 Q9PWK5	Q9PWK5 brachydanio
19	33	71.7	182	13 Q9DGP6	Q9DGP6 alepocephal

ALIGNMENTS

RESULT 1

Q9L3L1 PRELIMINARY; PRT; 487 AA.
AC Q9L3L1;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE PHOQ PROTEIN.
GN PHOQ.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SL1344;
RA Cano D.A., Martinez-Moya M., Casadesus J., Groisman E.A.,
Garcia-del Portillo F.;
RT "Attenuation of Salmonella proliferation within host cells mediated by
pathogen virulence regulators.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE
KINASES.
DR EMBL; AJ272210; CAB75592.1; -.
DR InterPro; IPR000410; -.
DR InterPro; IPR000658; -.
DR InterPro; IPR003594; -.
DR Pfam; PF00512; signal; 1.
DR Pfam; PF00672; DUF5; 1.
DR PRINTS; PR00344; BCTRLSENSOR.
DR SMART; SM00387; HATPase_c; 1.
KW Kinase; Phosphorylation; Sensory transduction; Transferase.
SQ SEQUENCE 487 AA; 55466 MW; BDCFEFC56F4CA058 CRC64;

Query Match 82.6%; Score 38; DB 2; Length 487;
Best Local Similarity 88.9%; Pred. No. 24;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GSGVLLSRK 10
|||||||

DB 329 GSGVLLSRK 337

Q9dgp4 lophius pls
Q9dgp2 rana pipien
Q9dgp1 chelydra se
Q9dgp0 chelydra se
Q9de59 xenopus lae
Q9k7ul bacillus ha
Q9slp8 citrus nats
Q93276 brachydanio
Q9ddl1 gallus gall
Q93124 treponema p
Q9ud87 homo sapien
Q9kuf6 vibrio chol
Q98128 vibrio para
Q9ug15 homo sapien
Q94800 homo sapien
Q9vzi7 drosophila
Q9vh75 drosophila
Q9vzi7 drosophila
Q95179 campylobact
Q95183 campylobact
Q9rf26 campylobact
Q9rf25 campylobact
Q9r953 campylobact
Q9r950 campylobact
Q9lat3 carassius a
Q9dgs6 carassius a
Q9y158 gallus gall

```

RESULT 2
Q9VR84
ID AC Q9VR84 PRELIMINARY; PRT; 709 AA.
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE CG5427 PROTEIN.
GN CG5427
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou P., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003636; AAF53195.1;
DR FlyBase; FBgn0032433; CG5427.
SQ SEQUENCE 709 AA; 79073 MW; 3404D1D8CBEBEDDD CRC64;

Query Match 78.3%; Score 36; DB 5; Length 709;
Best Local Similarity 60.0%; Pred. No. 91;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGSVLLSRK 10
Db 381 IGSVMVLSKK 390
:|||||:|

RESULT 3
Q9W737
ID AC Q9W737 PRELIMINARY; PRT; 1194 AA.
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE NOTCH-1 (FRAGMENT).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE=93328644; PubMed=10402194;
RA Wakamatsu Y., Maynard T.M., Jones S.U., Weston J.A.;
RT "NUMB localizes in the basal cortex of mitotic avian neuroepithelial
cells and modulates neuronal differentiation by binding to NOTCH-1.";
RL Neuron 23:71-81(1999).
DR EMBL; AF159231; AAD42893.1;
DR HSP; P00740; IEDM.
DR InterPro; IPR000561;
DR InterPro; IPR000800;
DR InterPro; IPR001064;
DR InterPro; IPR002110;
DR InterPro; IPR003571;
DR Pfam; PF000023; ank; 6.
DR Pfam; PF000066; notch; 3.
DR ProDom; PD000206;
DR PROSITE; PS30088; ANK_REPEAT; 4.
DR PROSITE; PS0297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR SMART; SM00248; ANK; 1.
FT NON_TER 1
SQ SEQUENCE 1194 AA; 130640 MW; 650380B8B6584974 CRC64;

Query Match 78.3%; Score 36; DB 13; Length 1194;
Best Local Similarity 70.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VGSVLLSRK 10
Db 392 IGSVLLSRK 401
:|||||:|

RESULT 4
Q9FF80
ID AC Q9FF80 PRELIMINARY; PRT; 670 AA.
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE SET-DOMAIN PROTEIN-LIKE.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=97471969; PubMed=9330910;
RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,
RA Miyajima N., Tabata S.,
RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence
features of the 1.6 Mb regions covered by twenty physically assigned
P1 clones.";
RL DNA Res. 4:215-230(1997).
DR EMBL; AB005245; BAB11516.1;
SQ SEQUENCE 670 AA; 74471 MW; C92CE89FF5C630F1 CRC64;

```

Query Match 76.1%; Score 35; DB 10; Length 670;
 Best Local Similarity 66.7%; Pred. No. 1.4e+02;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 GSGVLLSRK 10
 ||:|:|:|:|
 Db 461 GNGILVSRK 469

RESULT 5
 Q9F0D7 PRELIMINARY; PRT; 1505 AA.
 AC Q9F0D7;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE THAXTOMIN SYNTHETASE B.
 GN TXTB.
 OS Streptomyces acidiscabies.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=42234;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=84.104;
 RX MEDLINE=20566795; PubMed=11115114;
 RA Healy F.G., Wach M., Krasnoff S.B., Gibson D.M., Loria R.;
 RT "The txTAB genes of the plant pathogen Streptomyces acidiscabies
 RT encode a peptide synthetase required for phytoalexin thaxtomoin A
 RT production and pathogenicity.";
 RL Mol. Microbiol. 38:794-804(2000).
 DR EMBL; AF255732; AAG27088.1;
 SQ SEQUENCE 1505 AA; 163621 MW; 701A2038BD2B73EE CRC64;

Query Match 76.1%; Score 35; DB 2; Length 1505;
 Best Local Similarity 77.8%; Pred. No. 3.2e+02;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 VCGSVLLSR 9
 ||:|:|:|:|
 Db 570 VGTGLLSR 578

RESULT 6
 Q9L8H4 PRELIMINARY; PRT; 4247 AA.
 AC Q9L8H4;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE ACTINOMYCIN SYNTHETASE III.
 GN ACOM.
 OS Streptomyces chrysomallus.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1899;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 11523;
 RA Schaeffer F., Pfennig F., Grammel N., Keller U.;
 RT "Construction and in vitro analysis of a new bi-modular peptide
 RT synthetase for synthesis of N-methylated acyl-peptides.";
 RL Chem. Biol. 17:0-0(2000).
 DR EMBL; AF204401; AAF42473.1;
 DR InterPro; IPR000051;
 DR InterPro; IPR000255;
 DR InterPro; IPR000379;
 DR InterPro; IPR000873;
 DR InterPro; IPR001031;
 DR InterPro; IPR001242;

DR InterPro; IPR001601;
 DR Pfam; PF00501; AMP-binding; 3.
 DR Pfam; PF00550; pp-binding; 3.
 DR Pfam; PF00668; Condensation; 3.
 DR Pfam; PF00975; Thioesterase; 1.
 DR PRINTS; PR00154; AMPBINDING.
 DR PROSITE; PS00075; ACP_DOMAIN; 3.
 DR PROSITE; PS00455; AMP_BINDING; 3.
 DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; UNKNOWN_3.
 KW Phosphopantetheine.
 SQ SEQUENCE 4247 AA; 462299 MW; 580A7F41522A0BC5 CRC64;

Query Match 76.1%; Score 35; DB 2; Length 4247;
 Best Local Similarity 77.8%; Pred. No. 9.4e+02;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 VCGSVLLSR 9
 ||:|:|:|:|
 Db 2048 VGTGLLSR 2056

RESULT 7
 Q23577 PRELIMINARY; PRT; 80 AA.
 ID Q23577;
 AC Q23577;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
 DE COSMID ZK682.
 GN ZK682.3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifkin L., Roope A., Saunders D., Showkeen R.,
 RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Du Z., Le T.;
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Waterston R.;
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U41110; AAA82416.1;
 SQ SEQUENCE 80 AA; 9612 MW; 9A1766FD56C75E5C CRC64;

Query Match 73.9%; Score 34; DB 5; Length 80;
 Best Local Similarity 70.0%; Pred. No. 23;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VCGSVLLSRK 10
 ||:|:|:|:|
 Db 22 VGVNGVLLSR 31

RESULT 8

Q9HSG4
ID Q9HSG4 PRELIMINARY; PRT; 204 AA.
AC Q9HSG4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE VNG0244H.
GN VNG0244H.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leithauer B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudis J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.,
RT "Genome sequence of Halobacterium species NRC-1.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL: AE004988; AAG18842.1; B9995DBC86C713C6 CRC64;
SQ SEQUENCE 204 AA; 21180 MW; B9995DBC86C713C6 CRC64;

Query Match 73.9%; Score 34; DB 1; Length 204;
Best Local Similarity 70.0%; Pred. No. 61;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 VSGVLLSRK 10
Db 40 IGVGVLLSNK 49
:|:|:|:|:|
:|:|:|:|:|

RESULT 9
Q9V124 PRELIMINARY; PRT; 366 AA.
ID Q9V124;
AC Q9V124;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE HYPOTHETICAL 41.2 KDA PROTEIN.
GN PAB0416.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORSAY;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: Insights into archaeal chromosome
structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: A2248284; CAB49527.1;
KW Hypothetical protein.
SQ SEQUENCE 366 AA; 41214 MW; A2A6B78F0E3EF69C CRC64;

Query Match 73.9%; Score 34; DB 1; Length 366;
Best Local Similarity 66.7%; Pred. No. 1.1e+02;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 GSGVLLSRK 10
Db 237 GNGVLITRK 245
:|:|:|:|:|
:|:|:|:|:|

RESULT 10
Q9UAF9 PRELIMINARY; PRT; 528 AA.
ID Q9UAF9;
AC Q9UAF9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE SALK-7.
OS Ephydatia fluviatilis.
OC Eukaryota; Metazoa; Porifera; Demospongiae; Ceractinomorpha;
OC Haplosclerida; Spongillidae; Ephydatia.
OX NCBI_TaxID=31330;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99332084; PubMed=10405173;
RA Suga H., Ooc K., Miyata T.;
RT "Multiple TGF-beta receptor related genes in sponge and ancient gene
duplication before the parazoan-eumetazoan split.";
RL FEBS Lett. 453:346-350(1999).
CC -1- SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL: AB026830; BAA82607.1;
DR InterPro: IPR000333;
DR InterPro: IPR000561;
DR InterPro: IPR000719;
DR InterPro: IPR002290;
DR Pfam: PF00069; pkinase; 1.
DR PRINTS: PR00653; ACTIVIN2R.
DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR SMART: SM00220; S_TKc; 1.
DR ATP-binding; Serine/threonine-protein kinase; Transferase.
KW SEQUENCE 528 AA; 59299 MW; B864A0A63BA2671D CRC64;
SQ SEQUENCE 528 AA; 59299 MW; B864A0A63BA2671D CRC64;

Query Match 73.9%; Score 34; DB 5; Length 528;
Best Local Similarity 50.0%; Pred. No. 1.7e+02;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 VSGVLLSRK 10
Db 182 IGAGIMLRK 191
:|:|:|:|:|
:|:|:|:|:|

RESULT 11
Q9X0T3 PRELIMINARY; PRT; 82 AA.
ID Q9X0T3;
AC Q9X0T3;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE CONSERVED HYPOTHETICAL PROTEIN.
GN TM1206.
OS Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Steward A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
genome sequence of Thermotoga maritima.";
RL Nature 393:323-329(1999).
DR EMBL: AE001778; AAD36281.1;
DR TIGR: TM1206;
SQ SEQUENCE 82 AA; 9090 MW; 55D2B15BB48E8BFD CRC64;

Query Match 71.7%; Score 33; DB 2; Length 82;
Best Local Similarity 87.5%; Pred. No. 37;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VCGSVLLS 8
| | | | |
DB 9 VGAGVLLS 16

RESULT 12
Q9W663 PRELIMINARY; PRT; 182 AA.
AC Q9W663;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
DE GLUTAMATE DECARBOXYLASE ISOFORM 67 (EC 4.1.1.15) (FRAGMENT).
GN GAD67.
OS Trachemys scripta (Red-eared slider turtle) (Pseudemys scripta).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Testudines; Cryptodira; Testudinoidea; Emydidae; Trachemys.
OX NCBI_TaxID=34903;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=99261650; PubMed=10331265;
RA Bosma P.T., Blazquez M., Collins M.A., Bishop J.D.D., Drouin G.,
RA Priede I.G., Docherty K., Trudeau V.L.;
RT "Multiplicity of glutamic acid decarboxylases (GAD) in vertebrates:
molecular phylogeny and evidence for a new GAD paralog.";
RL Mol. Biol. Evol. 16:397-404(1999).
DR EMBL; AF043274; AAD22720.1; -.
DR InterPro; IPR000730; -.
DR Pfam; PF00282; pyridoxal_dec; 1.
DR ProDom; PD002673; -.
KW Lyase.
FT NON_TER 182
FT SEQUENCE 182 AA; 20047 MW; 27433BEEDE6218926 CRC64;

Query Match 71.7%; Score 33; DB 13; Length 182;
Best Local Similarity 66.7%; Pred. No. 86;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSGVLLSRK 10
| | | | |
DB 160 GGGLLSRK 168

RESULT 13
Q9W661 PRELIMINARY; PRT; 182 AA.
AC Q9W661;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
DE GLUTAMATE DECARBOXYLASE ISOFORM 65 (EC 4.1.1.15) (FRAGMENT).
GN GAD65.
OS Trachemys scripta (Red-eared slider turtle) (Pseudemys scripta).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Testudines; Cryptodira; Testudinoidea; Emydidae; Trachemys.
OX NCBI_TaxID=34903;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=99261650; PubMed=10331265;
RA Bosma P.T., Blazquez M., Collins M.A., Bishop J.D.D., Drouin G.,
RA Priede I.G., Docherty K., Trudeau V.L.;
RT "Multiplicity of glutamic acid decarboxylases (GAD) in vertebrates:
molecular phylogeny and evidence for a new GAD paralog.";
RL Mol. Biol. Evol. 16:397-404(1999).

DR EMBL; AF043272; AAD22718.1; -.
DR InterPro; IPR002129; -.
DR Pfam; PF00282; pyridoxal_dec; 1.
KW Lyase.
FT NON_TER 182
FT SEQUENCE 182 AA; 19942 MW; 7E0FEFC36DD0FB38 CRC64;

Query Match 71.7%; Score 33; DB 13; Length 182;
Best Local Similarity 66.7%; Pred. No. 86;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSGVLLSRK 10
| | | | |
DB 160 GGGLLSRK 168

RESULT 14
Q9W660 PRELIMINARY; PRT; 182 AA.
AC Q9W660;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
DE GLUTAMATE DECARBOXYLASE ISOFORM 67 (EC 4.1.1.15) (FRAGMENT).
GN GAD67.
OS Poephila guttata (Zebra finch) (Taeniopygia guttata).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Passeriformes; Estrildidae;
OC Estrildinae; Taeniopygia.
OX NCBI_TaxID=59729;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=99261650; PubMed=10331265;
RA Bosma P.T., Blazquez M., Collins M.A., Bishop J.D.D., Drouin G.,
RA Priede I.G., Docherty K., Trudeau V.L.;
RT "Multiplicity of glutamic acid decarboxylases (GAD) in vertebrates:
molecular phylogeny and evidence for a new GAD paralog.";
RL Mol. Biol. Evol. 16:397-404(1999).
DR EMBL; AF043271; AAD22717.1; -.
DR InterPro; IPR002129; -.
DR Pfam; PF00282; pyridoxal_dec; 1.
KW Lyase.
FT NON_TER 182
FT SEQUENCE 182 AA; 20057 MW; 26FFF4A0B3F1F9E9 CRC64;

Query Match 71.7%; Score 33; DB 13; Length 182;
Best Local Similarity 66.7%; Pred. No. 86;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSGVLLSRK 10
| | | | |
DB 160 GGGLLSRK 168

RESULT 15
Q9W659 PRELIMINARY; PRT; 182 AA.
AC Q9W659;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
DE GLUTAMATE DECARBOXYLASE ISOFORM 65 (EC 4.1.1.15) (FRAGMENT).
GN GAD65.
OS Poephila guttata (Zebra finch) (Taeniopygia guttata).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Passeriformes; Estrildidae;
OC Estrildinae; Taeniopygia.
OX NCBI_TaxID=59729;

Fri Sep 7 10:58:27 2001

```

RN SEQUENCE FROM N.A.
RC TISSUE-BRAIN:
RA MEDLINE-99261650; PubMed-10331265;
RA Bosma P.T., Blazquez M., Collins M.A., Bishop J.D.D., Drouin G.,
RA Priede I.G., Docherty K., Trudeau V.L.;
RA "Multiplicy of glutamic acid decarboxylases (GAD) in vertebrates:
RA molecular phylogeny and evidence for a new GAD paralog.";
RL Mol. Biol. Evol. 16:397-404(1999).
DR EMBL; AF043270; AAD22716.1; -
DR InterPro; IPR002129; -
DR Pfam; PF00282; pyridoxal_deC; 1.
KW Lyase.
FT NON_TER 1 1
FT NON_TER 182 182
SQ SEQUENCE 182 AA; 19918 MW; A35335D6DC87122A CRC64;

Query Match 71.7%; Score 33; DB 13; Length 182;
Best Local Similarity 66.7%; Pred. No. 86;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSGVLLSRK 10
   |::|||
Db 160 GGGLLSRK 168

```

Search completed: September 6, 2001, 16:49:58
Job time: 742 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:51:12 ; Search time 72.75 Seconds
(without alignments)
4.709 Million cell updates/sec

Title: US-09-603-713-30
Perfect score: 46
Sequence: 1 VSGVLLSRK 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	89.1	2444	1 NTCL_HUMAN	P46531 homo sapien
2	41	89.1	2531	1 NTCL_MOUSE	Q01705 mus musculus
3	41	89.1	2531	1 NTCL_RAT	Q07008 rattus norv
4	38	82.6	451	1 PHOQ_SALTY	P14147 salmoneilla
5	33	71.7	258	1 LLDR_ECOLI	P33233 escherichia
6	33	71.7	575	1 FLA2_CAMJE	P22251 campylobact
7	33	71.7	575	1 FLB2_CAMJE	P22252 campylobact
8	33	71.7	585	1 DCE2_HUMAN	Q05329 homo sapien
9	33	71.7	585	1 DCE2_MOUSE	P48320 mus musculus
10	33	71.7	585	1 DCE2_PIG	P48321 sus scrofa
11	33	71.7	585	1 DCE2_RAT	Q05683 rattus norv
12	33	71.7	593	1 DCE1_MOUSE	P48318 mus musculus
13	33	71.7	593	1 DCE1_RAT	P18088 rattus norv
14	33	71.7	594	1 DCE1_FELCA	P14748 felis silve
15	33	71.7	594	1 DCE1_HUMAN	Q99259 homo sapien
16	33	71.7	594	1 DCE1_PIG	P48319 sus scrofa
17	32	69.6	660	1 SGAT_MYCPN	P75291 mycoplasma
18	31	67.4	117	1 ARRI_ECOLI	P15905 escherichia
19	31	67.4	137	1 Y4JC_RHISN	P55503 rhizobium s
20	31	67.4	137	1 PFDA_ARCFU	Q28216 archaeoglob
21	31	67.4	296	1 YEM4_SCHPO	O14027 schizosacch
22	31	67.4	333	1 YC48_CYAPA	P48325 cyanophora
23	31	67.4	340	1 DEGS_HAEIN	P44947 haemophilus
24	31	67.4	347	1 WZEE_SALTY	O33789 salmoneilla
25	31	67.4	349	1 DMC1_LITLO	P37384 lilium long
26	31	67.4	349	1 WZEE_ECOLI	P25905 escherichia
27	31	67.4	373	1 DDL_MYCSM	Q929n0 mycobacteri
28	31	67.4	383	1 XYLR_STAXY	P27159 staphylococ
29	31	67.4	397	1 PURA_THEMEA	Q9x011 thermotoga
30	31	67.4	426	1 Y680_CHLPN	P927m4 chlamydia p
31	31	67.4	487	1 BAT1_MOUSE	Q9qxa6 mus musculus
32	31	67.4	487	1 BAT1_RAT	P82252 rattus norv
33	31	67.4	496	1 C4AE_DROME	O46054 drosophila

RESULT 1
NTCL_HUMAN
ID NTCL_HUMAN STANDARD: PRT: 2444 AA.
AC P46531;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 1 PRECURSOR (TRANSLOCATION-
DE ASSOCIATED NOTCH PROTEIN TAN-1) (FRAGMENT).
GN NOTCH1 OR TANI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91347367; PubMed=1831692;
RA Ellisen L.W., Bird J., West D.C., Soreng A.L., Reynolds T.C.,
RA Smith S.D., Sklar J.;
RT "TAN-1, the human homolog of the Drosophila notch gene, is broken by
RT chromosomal translocations in T lymphoblastic neoplasms.";
RL Cell 66:649-661(1991).
CC -!- FUNCTION: MAY BE IMPORTANT FOR NORMAL LYMPHOCYTE FUNCTION. IN
CC ALTERED FORM, MAY CONTRIBUTE TO TRANSFORMATION OR PROGRESSION
CC IN SOME T-CELL NEOPLASMS.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: IN FETAL TISSUES MOST ABUNDANT IN SPLEEN,
CC BRAIN STEM AND LUNG. ALSO PRESENT IN MOST ADULT TISSUES WHERE IT
CC IS FOUND MAINLY IN LYMPHOID TISSUES.
CC -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -!- SIMILARITY: CONTAINS 5 ANK REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M7398C; AAA60614.1; -
CC HSP: P0074C; 11XA.
CC MIM: 190198; -
CC InterPro: IPR000152; -
CC InterPro: IPR0000561; -
CC InterPro: IPR000800; -
CC InterPro: IPR001881; -
CC InterPro: IPR002110; -
CC Pfam: PF00008; EGF; 36.
CC Pfam: PF00023; ank; 6.
CC PROSITE: PS50088; ANK_REPEAT; 4.
CC PROSITE: PS50297; ANK_REPEAT; 1.
CC PROSITE: PS50010; ASX_HYDROXYL; 20.

ALIGNMENTS

34	31	67.4	502	1	XCT_MOUSE	Q9wtr6 mus musculus
35	31	67.4	533	1	MUTL_THEAO	P96082 thermus aqu
36	31	67.4	800	1	ARNT_RAT	P41739 rattus norv
37	31	67.4	2199	1	DPOE_SCHPO	P87154 schizosacch
38	30	65.2	154	1	GLB2_NIPBR	P51535 nipostroch
39	30	65.2	163	1	LSPA_HELPU	Q9zm23 helicobacte
40	30	65.2	171	1	I17A_HUMAN	Q95955 homo sapien
41	30	65.2	171	1	I17A_MOUSE	Q920V8 mus musculus
42	30	65.2	171	1	I17A_RAT	O35092 rattus norv
43	30	65.2	171	1	YR18_THEPE	P15890 thermofilum
44	30	65.2	233	1	YCBV_PSEDE	P29938 pseudomonas
45	30	65.2	238	1	YRN7_CAEEL	Q09421 caenorhabdi

DR PROSITE; PS00022; EGF_1; 34.
DR PROSITE; PS01186; EGF_2; 26.
DR PROSITE; PS01187; EGF_CA; 18.
KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Signal; Glycoprotein.
FT SIGNAL 1 18
FT CHAIN 19 >2444
FT DOMAIN 19 1736
FT TRANSMEM 1737 1757
FT DOMAIN 1758 >2444
FT DOMAIN 20 58
FT DOMAIN 59 99
FT DOMAIN 102 139
FT DOMAIN 140 176
FT DOMAIN 178 216
FT DOMAIN 218 255
FT DOMAIN 257 293
FT DOMAIN 295 333
FT DOMAIN 335 371
FT DOMAIN 372 410
FT DOMAIN 412 450
FT DOMAIN 452 488
FT DOMAIN 490 526
FT DOMAIN 528 564
FT DOMAIN 566 601
FT DOMAIN 603 639
FT DOMAIN 641 676
FT DOMAIN 678 714
FT DOMAIN 716 751
FT DOMAIN 753 789
FT DOMAIN 791 827
FT DOMAIN 829 868
FT DOMAIN 870 906
FT DOMAIN 908 944
FT DOMAIN 946 982
FT DOMAIN 984 1020
FT DOMAIN 1022 1058
FT DOMAIN 1060 1096
FT DOMAIN 1098 1144
FT DOMAIN 1146 1182
FT DOMAIN 1184 1220
FT DOMAIN 1222 1266
FT DOMAIN 1268 1306
FT DOMAIN 1308 1347
FT DOMAIN 1349 1385
FT DOMAIN 1388 1427
FT REPEAT 1446 1481
FT REPEAT 1482 1523
FT REPEAT 1524 1563
FT REPEAT 1563 1597
FT REPEAT 1597 1641
FT REPEAT 1641 1685
FT REPEAT 1685 1729
FT REPEAT 1729 1773
FT REPEAT 1773 1817
FT REPEAT 1817 1861
FT REPEAT 1861 1905
FT REPEAT 1905 1949
FT REPEAT 1949 1993
FT REPEAT 1993 2037
FT REPEAT 2037 2081
FT REPEAT 2081 2125
FT REPEAT 2125 2169
FT REPEAT 2169 2213
FT REPEAT 2213 2257
FT REPEAT 2257 2301
FT REPEAT 2301 2345
FT REPEAT 2345 2389
FT REPEAT 2389 2433
FT REPEAT 2433 2477
FT REPEAT 2477 2521
FT REPEAT 2521 2565
FT REPEAT 2565 2609
FT REPEAT 2609 2653
FT REPEAT 2653 2697
FT REPEAT 2697 2741
FT REPEAT 2741 2785
FT REPEAT 2785 2829
FT REPEAT 2829 2873
FT REPEAT 2873 2917
FT REPEAT 2917 2961
FT REPEAT 2961 3005
FT REPEAT 3005 3049
FT REPEAT 3049 3093
FT REPEAT 3093 3137
FT REPEAT 3137 3181
FT REPEAT 3181 3225
FT REPEAT 3225 3269
FT REPEAT 3269 3313
FT REPEAT 3313 3357
FT REPEAT 3357 3401
FT REPEAT 3401 3445
FT REPEAT 3445 3489
FT REPEAT 3489 3533
FT REPEAT 3533 3577
FT REPEAT 3577 3621
FT REPEAT 3621 3665
FT REPEAT 3665 3709
FT REPEAT 3709 3753
FT REPEAT 3753 3797
FT REPEAT 3797 3841
FT REPEAT 3841 3885
FT REPEAT 3885 3929
FT REPEAT 3929 3973
FT REPEAT 3973 4017
FT REPEAT 4017 4061
FT REPEAT 4061 4105
FT REPEAT 4105 4149
FT REPEAT 4149 4193
FT REPEAT 4193 4237
FT REPEAT 4237 4281
FT REPEAT 4281 4325
FT REPEAT 4325 4369
FT REPEAT 4369 4413
FT REPEAT 4413 4457
FT REPEAT 4457 4501
FT REPEAT 4501 4545
FT REPEAT 4545 4589
FT REPEAT 4589 4633
FT REPEAT 4633 4677
FT REPEAT 4677 4721
FT REPEAT 4721 4765
FT REPEAT 4765 4809
FT REPEAT 4809 4853
FT REPEAT 4853 4897
FT REPEAT 4897 4941
FT REPEAT 4941 4985
FT REPEAT 4985 5029
FT REPEAT 5029 5073
FT REPEAT 5073 5117
FT REPEAT 5117 5161
FT REPEAT 5161 5205
FT REPEAT 5205 5249
FT REPEAT 5249 5293
FT REPEAT 5293 5337
FT REPEAT 5337 5381
FT REPEAT 5381 5425
FT REPEAT 5425 5469
FT REPEAT 5469 5513
FT REPEAT 5513 5557
FT REPEAT 5557 5601
FT REPEAT 5601 5645
FT REPEAT 5645 5689
FT REPEAT 5689 5733
FT REPEAT 5733 5777
FT REPEAT 5777 5821
FT REPEAT 5821 5865
FT REPEAT 5865 5909
FT REPEAT 5909 5953
FT REPEAT 5953 5997
FT REPEAT 5997 6041
FT REPEAT 6041 6085
FT REPEAT 6085 6129
FT REPEAT 6129 6173
FT REPEAT 6173 6217
FT REPEAT 6217 6261
FT REPEAT 6261 6305
FT REPEAT 6305 6349
FT REPEAT 6349 6393
FT REPEAT 6393 6437
FT REPEAT 6437 6481
FT REPEAT 6481 6525
FT REPEAT 6525 6569
FT REPEAT 6569 6613
FT REPEAT 6613 6657
FT REPEAT 6657 6701
FT REPEAT 6701 6745
FT REPEAT 6745 6789
FT REPEAT 6789 6833
FT REPEAT 6833 6877
FT REPEAT 6877 6921
FT REPEAT 6921 6965
FT REPEAT 6965 7009
FT REPEAT 7009 7053
FT REPEAT 7053 7097
FT REPEAT 7097 7141
FT REPEAT 7141 7185
FT REPEAT 7185 7229
FT REPEAT 7229 7273
FT REPEAT 7273 7317
FT REPEAT 7317 7361
FT REPEAT 7361 7405
FT REPEAT 7405 7449
FT REPEAT 7449 7493
FT REPEAT 7493 7537
FT REPEAT 7537 7581
FT REPEAT 7581 7625
FT REPEAT 7625 7669
FT REPEAT 7669 7713
FT REPEAT 7713 7757
FT REPEAT 7757 7801
FT REPEAT 7801 7845
FT REPEAT 7845 7889
FT REPEAT 7889 7933
FT REPEAT 7933 7977
FT REPEAT 7977 8021
FT REPEAT 8021 8065
FT REPEAT 8065 8109
FT REPEAT 8109 8153
FT REPEAT 8153 8197
FT REPEAT 8197 8241
FT REPEAT 8241 8285
FT REPEAT 8285 8329
FT REPEAT 8329 8373
FT REPEAT 8373 8417
FT REPEAT 8417 8461
FT REPEAT 8461 8505
FT REPEAT 8505 8549
FT REPEAT 8549 8593
FT REPEAT 8593 8637
FT REPEAT 8637 8681
FT REPEAT 8681 8725
FT REPEAT 8725 8769
FT REPEAT 8769 8813
FT REPEAT 8813 8857
FT REPEAT 8857 8901
FT REPEAT 8901 8945
FT REPEAT 8945 8989
FT REPEAT 8989 9033
FT REPEAT 9033 9077
FT REPEAT 9077 9121
FT REPEAT 9121 9165
FT REPEAT 9165 9209
FT REPEAT 9209 9253
FT REPEAT 9253 9297
FT REPEAT 9297 9341
FT REPEAT 9341 9385
FT REPEAT 9385 9429
FT REPEAT 9429 9473
FT REPEAT 9473 9517
FT REPEAT 9517 9561
FT REPEAT 9561 9605
FT REPEAT 9605 9649
FT REPEAT 9649 9693
FT REPEAT 9693 9737
FT REPEAT 9737 9781
FT REPEAT 9781 9825
FT REPEAT 9825 9869
FT REPEAT 9869 9913
FT REPEAT 9913 9957
FT REPEAT 9957 10001
FT REPEAT 10001 10045
FT REPEAT 10045 10089
FT REPEAT 10089 10133
FT REPEAT 10133 10177
FT REPEAT 10177 10221
FT REPEAT 10221 10265
FT REPEAT 10265 10309
FT REPEAT 10309 10353
FT REPEAT 10353 10397
FT REPEAT 10397 10441
FT REPEAT 10441 10485
FT REPEAT 10485 10529
FT REPEAT 10529 10573
FT REPEAT 10573 10617
FT REPEAT 10617 10661
FT REPEAT 10661 10705
FT REPEAT 10705 10749
FT REPEAT 10749 10793
FT REPEAT 10793 10837
FT REPEAT 10837 10881
FT REPEAT 10881 10925
FT REPEAT 10925 10969
FT REPEAT 10969 11013
FT REPEAT 11013 11057
FT REPEAT 11057 11101
FT REPEAT 11101 11145
FT REPEAT 11145 11189
FT REPEAT 11189 11233
FT REPEAT 11233 11277
FT REPEAT 11277 11321
FT REPEAT 11321 11365
FT REPEAT 11365 11409
FT REPEAT 11409 11453
FT REPEAT 11453 11497
FT REPEAT 11497 11541
FT REPEAT 11541 11585
FT REPEAT 11585 11629
FT REPEAT 11629 11673
FT REPEAT 11673 11717
FT REPEAT 11717 11761
FT REPEAT 11761 11805
FT REPEAT 11805 11849
FT REPEAT 11849 11893
FT REPEAT 11893 11937
FT REPEAT 11937 11981
FT REPEAT 11981 12025
FT REPEAT 12025 12069
FT REPEAT 12069 12113
FT REPEAT 12113 12157
FT REPEAT 12157 12201
FT REPEAT 12201 12245
FT REPEAT 12245 12289
FT REPEAT 12289 12333
FT REPEAT 12333 12377
FT REPEAT 12377 12421
FT REPEAT 12421 12465
FT REPEAT 12465 12509
FT REPEAT 12509 12553
FT REPEAT 12553 12597
FT REPEAT 12597 12641
FT REPEAT 12641 12685
FT REPEAT 12685 12729
FT REPEAT 12729 12773
FT REPEAT 12773 12817
FT REPEAT 12817 12861
FT REPEAT 12861 12905
FT REPEAT 12905 12949
FT REPEAT 12949 12993
FT REPEAT 12993 13037
FT REPEAT 13037 13081
FT REPEAT 13081 13125
FT REPEAT 13125 13169
FT REPEAT 13169 13213
FT REPEAT 13213 13257
FT REPEAT 13257 13301
FT REPEAT 13301 13345
FT REPEAT 13345 13389
FT REPEAT 13389 13433
FT REPEAT 13433 13477
FT REPEAT 13477 13521
FT REPEAT 13521 13565
FT REPEAT 13565 13609
FT REPEAT 13609 13653
FT REPEAT 13653 13697
FT REPEAT 13697 13741
FT REPEAT 13741 13785
FT REPEAT 13785 13829
FT REPEAT 13829 13873
FT REPEAT 13873 13917
FT REPEAT 13917 13961
FT REPEAT 13961 14005
FT REPEAT 14005 14049
FT REPEAT 14049 14093
FT REPEAT 14093 14137
FT REPEAT 14137 14181
FT REPEAT 14181 14225
FT REPEAT 14225 14269
FT REPEAT 14269 14313
FT REPEAT 14313 14357
FT REPEAT 14357 14401
FT REPEAT 14401 14445
FT REPEAT 14445 14489
FT REPEAT 14489 14533
FT REPEAT 14533 14577
FT REPEAT 14577 14621
FT REPEAT 14621 14665
FT REPEAT 14665 14709
FT REPEAT 14709 14753
FT REPEAT 14753 14797
FT REPEAT 14797 14841
FT REPEAT 14841 14885
FT REPEAT 14885 14929
FT REPEAT 14929 14973
FT REPEAT 14973 15017
FT REPEAT 15017 15061
FT REPEAT 15061 15105
FT REPEAT 15105 15149
FT REPEAT 15149 15193
FT REPEAT 15193 15237
FT REPEAT 15237 15281
FT REPEAT 15281 15325
FT REPEAT 15325 15369
FT REPEAT 15369 15413
FT REPEAT 15413 15457
FT REPEAT 15457 15501
FT REPEAT 15501 15545
FT REPEAT 15545 15589
FT REPEAT 15589 15633
FT REPEAT 15633 15677
FT REPEAT 15677 15721
FT REPEAT 15721 15765
FT REPEAT 15765 15809
FT REPEAT 15809 15853
FT REPEAT 15853 15897
FT REPEAT 15897 15941
FT REPEAT 15941 15985
FT REPEAT 15985 16029
FT REPEAT 16029 16073
FT REPEAT 16073 16117
FT REPEAT 16117 16161
FT REPEAT 16161 16205
FT REPEAT 16205 16249
FT REPEAT 16249 16293
FT REPEAT 16293 16337
FT REPEAT 16337 16381
FT REPEAT 16381 16425
FT REPEAT 16425 16469
FT REPEAT 16469 16513
FT REPEAT 16513 16557
FT REPEAT 16557 16601
FT REPEAT 16601 16645
FT REPEAT 16645 16689
FT REPEAT 16689 16733
FT REPEAT 16733 16777
FT REPEAT 16777 16821
FT REPEAT 16821 16865
FT REPEAT 16865 16909
FT REPEAT 16909 16953
FT REPEAT 16953 16997
FT REPEAT 16997 17041
FT REPEAT 17041 17085
FT REPEAT 17085 17129
FT REPEAT 17129 17173
FT REPEAT 17173 17217
FT REPEAT 17217 17261
FT REPEAT 17261 17305
FT REPEAT 17305 17349
FT REPEAT 17349 17393
FT REPEAT 17393 17437
FT REPEAT 17437 17481
FT REPEAT 17481 17525
FT REPEAT 17525 17569
FT REPEAT 17569 17613
FT REPEAT 17613 17657
FT REPEAT 17657 17701
FT REPEAT 17701 17745
FT REPEAT 17745 17789
FT REPEAT 17789 17833
FT REPEAT 17833 17877
FT REPEAT 17877 17921
FT REPEAT 17921 17965
FT REPEAT 17965 18009
FT REPEAT 18009 18053
FT REPEAT 18053 18097
FT REPEAT 18097 18141
FT REPEAT 18141 18185
FT REPEAT 18185 18229
FT REPEAT 18229 18273
FT REPEAT 18273 18317
FT REPEAT 18317 18361
FT REPEAT 18361 18405
FT REPEAT 18405 18449
FT REPEAT 18449 18493
FT REPEAT 18493 18537
FT REPEAT 18537 18581
FT REPEAT 18581 18625
FT REPEAT 18625 18669
FT REPEAT 18669 18713
FT REPEAT 18713 18757
FT REPEAT 18757 18801
FT REPEAT 18801 18845
FT REPEAT 18845 18889
FT REPEAT 18889 18933
FT REPEAT 18933 18977
FT REPEAT 18977 19021
FT REPEAT 19021 19065
FT REPEAT 19065 19109
FT REPEAT 19109 19153
FT REPEAT 19153 19197
FT REPEAT 19197 19241
FT REPEAT 19241 19285
FT REPEAT 19285 19329
FT REPEAT 19329 19373
FT REPEAT 19373 19417
FT REPEAT 19417 19461
FT REPEAT 19461 19505
FT REPEAT 19505 19549
FT REPEAT 19549 19593
FT REPEAT 19593 19637
FT REPEAT 19637 19681
FT REPEAT 19681 19725
FT REPEAT 19725 19769
FT REPEAT 19769 19813
FT REPEAT 19813 19857
FT REPEAT 19857 19901
FT REPEAT 19901 19945
FT REPEAT 19945 19989
FT REPEAT 19989 20033
FT REPEAT 20033 20077
FT REPEAT 20077 20121
FT REPEAT 20121 20165
FT REPEAT 20165 20209
FT REPEAT 20209 20253
FT REPEAT 20253 20297
FT REPEAT 20297 20341
FT REPEAT 20341 20385
FT REPEAT 20385 20429
FT REPEAT 20429 20473
FT REPEAT 20473 20517
FT REPEAT 20517 20561
FT REPEAT 20561 20605
FT REPEAT 20605 20649
FT REPEAT 20649 20693
FT REPEAT 20693 20737
FT REPEAT 20737 20781
FT REPEAT 20781 20825
FT REPEAT 20825 20869
FT REPEAT 20869 20913
FT REPEAT 20913 20957
FT REPEAT 20957 21001
FT REPEAT 21001 21045
FT REPEAT 21045 21089
FT REPEAT 21089 21133
FT REPEAT 21133 21177
FT REPEAT 21177 21221
FT REPEAT 21221 21265
FT REPEAT 21265 21309
FT REPEAT 21309 21353
FT REPEAT 21353 21397
FT REPEAT 21397 21441
FT REPEAT 21441 21485
FT REPEAT 21485 21529
FT REPEAT 21529 21573
FT REPEAT 21573 21617
FT REPEAT 21617 21661
FT REPEAT 21661 21705
FT REPEAT 21705 21749
FT REPEAT 21749 21793
FT REPEAT 21793 21837
FT REPEAT 21837 21881
FT REPEAT 21881 21925
FT REPEAT 21925 21969
FT REPEAT 21969 22013
FT REPEAT 22013 22057
FT REPEAT 22057 22101
FT REPEAT 22101 22145
FT REPEAT 22145 22189
FT REPEAT 22189 22233
FT REPEAT 22233 22277
FT REPEAT 22277 22321
FT REPEAT 22321 22365
FT REPEAT 22365 22409
FT REPEAT 22409 22453
FT REPEAT 22453 22497
FT REPEAT 22497 22541
FT REPEAT 22541 22585
FT REPEAT 22585 22629
FT REPEAT 22629 22673
FT REPEAT 22673 22717
FT REPEAT 22717 22761
FT REPEAT 22761 22805
FT REPEAT 22805 22849
FT REPEAT 22849 22893
FT REPEAT 22893 22937
FT REPEAT 22937 22981
FT REPEAT 22981 23025
FT REPEAT 23025 23069
FT REPEAT 23069 23113
FT REPEAT 23113 23157
FT REPEAT 23157 23201
FT REPEAT 23201 23245
FT REPEAT 23245 23289
FT REPEAT 23289 23333
FT REPEAT 23333 23377
FT REPEAT 23377 23421
FT REPEAT 23421 23465
FT REPEAT 23465 23509
FT REPEAT 23509 23553
FT REPEAT 23553 23597
FT REPEAT 23597 23641
FT REPEAT 23641 23685
FT REPEAT 23685 23729
FT REPEAT 23729 23773
FT REPEAT 23773 23817
FT REPEAT 23817 23861
FT REPEAT 23861 23905
FT REPEAT 23905 23949
FT REPEAT 23949 23993
FT REPEAT 23993 24037
FT REPEAT 24037 24081
FT REPEAT 24081 24125
FT REPEAT 24125 24169
FT REPEAT 24169 24213
FT REPEAT 24213 24257
FT REPEAT 24257 24301
FT REPEAT 24301 24345
FT REPEAT 24345 24389
FT REPEAT 24389 24433
FT REPEAT 24433 24477
FT REPEAT 24477 24521
FT REPEAT 24521 24565
FT REPEAT 24565 24609
FT REPEAT 24609 24653
FT REPEAT 24653 24697
FT REPEAT 24697 24741
FT REPEAT 24741 24785
FT REPEAT 24785 24829
FT REPEAT 24829 24873
FT REPEAT 24873 24917
FT REPEAT 24917 24961
FT REPEAT 24961 25005
FT REPEAT 25005 25049
FT REPEAT 25049 25093
FT REPEAT 25093 25137
FT REPEAT 25137 25181
FT REPEAT 25181 25225
FT REPEAT 25225 25269
FT REPEAT 25269 25313
FT REPEAT 25313 25357
FT REPEAT 25357 25401
FT REPEAT 25401 25445
FT REPEAT 25445 25489
FT REPEAT 25489 25533
FT REPEAT 25533 25577
FT REPEAT 25577 25621
FT REPEAT 25621 25665
FT REPEAT 25665 25709
FT REPEAT 25709 25753
FT REPEAT 25753 25797
FT REPEAT 25797 25841
FT REPEAT 25841 25885
FT REPEAT 25885 25929
FT REPEAT 25929 25973
FT REPEAT 25973 26017
FT REPEAT 26017 26061
FT REPEAT 26061 26105
FT REPEAT 26105 26149
FT REPEAT 26149 26193
FT REPEAT 26193 26237
FT REPEAT 26237 26281
FT REPEAT 26281 26325
FT REPEAT 26325 26369
FT REPEAT 26369 26413
FT REPEAT 26413 26457
FT REPEAT 26457 26501
FT REPEAT 26501 26545
FT REPEAT 26545 26589
FT REPEAT 26589 26633
FT REPEAT 26633 26677
FT REPEAT 26677 26721
FT REPEAT 26721 26765
FT REPEAT 26765 26809
FT REPEAT 26809 26853
FT REPEAT 26853 26897
FT REPEAT 26897 26941
FT REPEAT 26941 26985
FT REPEAT 26985 27029
FT REPEAT 27029 27073
FT REPEAT 27073 27117
FT REPEAT 27117 27161
FT REPEAT 27161 27205
FT REPEAT 27205 27249
FT REPEAT 27249 27293
FT REPEAT 27293 27337
FT REPEAT 27337 27381
FT REPEAT 27381 27425
FT REPEAT 27425 27469
FT REPEAT 27469 27513
FT REPEAT 27513 27557
FT REPEAT 27557 27601
FT REPEAT 27601 27645
FT REPEAT 27645 27689
FT REPEAT 27689 27733
FT REPEAT 27733 27777
FT REPEAT 27777 27821
FT REPEAT 27821 27865
FT REPEAT 27865 27909
FT REPEAT 27909 27953
FT REPEAT 27953 27997
FT REPEAT 27997 28041
FT REPEAT 28041 28085
FT REPEAT 28085 28129
FT REPEAT 28129 28173
FT REPEAT 28173 28217
FT REPEAT 28217 28261
FT REPEAT 28261 28305
FT REPEAT 28305 28349
FT REPEAT 28349 28393
FT REPEAT 28393 28437
FT REPEAT 28437 28481
FT REPEAT 28481 28525
FT REPEAT 28525 28569
FT REPEAT 28569 28613
FT REPEAT 28613 28657
FT REPEAT 28657 28701
FT REPEAT 28701 28745
FT REPEAT 28745 28789
FT REPEAT 28789 28833
FT REPEAT 28833 28877
FT REPEAT 28877 28921
FT REPEAT 28921 28965
FT REPEAT 28965 29009
FT REPEAT 29009 29053
FT REPEAT 29053 29097
FT REPEAT 29097 29141
FT REPEAT 29141 29185
FT REPEAT 29185 29229
FT REPEAT 29229 29273
FT REPEAT 29273 29317
FT REPEAT 29317 29361
FT REPEAT 29361 29405
FT REPEAT 29405 29449
FT REPEAT 29449 29493
FT REPEAT 29493 29537
FT REPEAT 29537 29581
FT REPEAT 29581 29625
FT REPEAT 29625 29669
FT REPEAT 29669 29713
FT REPEAT 29713 29757
FT REPEAT 29757 29801
FT REPEAT 29801 29845
FT REPEAT 29845 29889
FT REPEAT 29889 29933
FT REPEAT 29933 29977
FT REPEAT 29977 30021
FT REPEAT 30021 30065
FT REPEAT 30065 30109
FT REPEAT 30109 30153
FT REPEAT 30153 30197
FT REPEAT 30197 30241
FT REPEAT 30241 30285
FT REPEAT 30285 30329
FT REPEAT 30329 30373
FT REPEAT 30373 30417
FT REPEAT 30417 30461
FT REPEAT 30461 30505
FT REPEAT 30505 30549
FT REPEAT 30549 30593
FT REPEAT 30593 30637
FT REPEAT 30637 30681
FT REPEAT 30681 30725
FT REPEAT 30725 30769
FT REPEAT 30769 30813
FT REPEAT 30813 30857
FT REPEAT 30857 30901
FT REPEAT 30901 30945
FT REPEAT 30945 30989
FT REPEAT 30989 31033
FT REPEAT 31033 31077
FT REPEAT 31077 31121
FT REPEAT 31121 31165
FT REPEAT 31165 31209
FT REPEAT 31209 31253
FT REPEAT 31253 31297
FT REPEAT 31297 31341
FT REPEAT 31341 31385
FT REPEAT 31385 31429
FT REPEAT 31429 31473
FT REPEAT 31473 31517
FT REPEAT 31517 31561
FT REPEAT 31561 31605
FT REPEAT 31605 31649
FT REPEAT 31649 31693
FT REPEAT 31693 31737
FT REPEAT 31737 31781
FT REPEAT 31781 31825
FT REPEAT 31825 31869
FT REPEAT 31869 31913
FT REPEAT 31913 31957
FT REPEAT 31957 32001
FT REPEAT 32001 32045
FT REPEAT 32045 32089
FT REPEAT 32089 32133
FT REPEAT 32133 32177
FT REPEAT 32177 32221
FT REPEAT 32221 32265
FT REPEAT 32265 32309
FT REPEAT 32309 32353
FT REPEAT 32353 32397
FT REPEAT 32397 32441
FT REPEAT 32441

Query Match 89.1%; Score 41; DB 1; Length 2444;
 Best Local Similarity 90.0%; Pred. No. 4.1;
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VCGSVLLSRK 10
 DB 1751 VCGSVLLSRK 1760
 || |||||
 AC Q01705; STANDARD; PRT; 2531 AA.
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR (NOTCH PROTEIN).
 GN NOTCH1 OR NOTCH.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=93194170; PubMed=8449489;
 RA Franco del Amo F., Gendron-Maguire M., Swiatek P.J., Jenkins N.A.,
 RA Copeland N.G., Gridley T.;
 RT "Cloning, analysis, and chromosomal localization of Notch-1, a mouse
 RT homolog of Drosophila Notch.";
 RT Genomics 15:259-264(1993).
 RN [2]
 RP SEQUENCE OF 1551-2170 FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=93048835; PubMed=1425352;
 RA Franco del Amo F., Smith D.E., Swiatek P.J., Gendron-Maguire M.,
 RA Greenspan R.J., McMahon A.P., Gridley T.;
 RT "Expression pattern of Notch, a mouse homolog of Drosophila Notch,
 RT suggests an important role in early postimplantation mouse
 RT development.";
 RL Development 115:737-744(1992).
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
 CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
 CC -1- SIMILARITY: CONTAINS 5 ANK REPEATS.
 CC -1- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; Z11886; CAA77941.1;
 CC HSP; P00740; 11XA.
 DR MGD; MGI:97363; Notch1.
 DR InterPro; IPR000152; -
 DR InterPro; IPR000561; -
 DR InterPro; IPR000800; -
 DR InterPro; IPR001438; -
 DR InterPro; IPR001881; -
 DR InterPro; IPR002110; -
 DR Pfam; PF00008; EGF; 35.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00066; notch; 3.
 DR PRINTS; PR00010; EGFBL00.
 DR PROSITE; PS50088; ANK_REPEAT; 2.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; 22.
 DR

DR PROSITE; PS00022; EGF_1; 34.
 DR PROSITE; PS01186; EGF_2; 27.
 DR PROSITE; PS01187; EGF_CA; 21.
 KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Signal; Glycoprotein.
 FT SIGNAL 1 18 POTENTIAL.
 FT CHAIN 19 2531 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
 FT DOMAIN 19 1725 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1726 1746 POTENTIAL.
 FT DOMAIN 1747 2531 CYTOPLASMIC (POTENTIAL).
 FT EGF-LIKE 1 58 EGF-LIKE 1.
 FT EGF-LIKE 2 59 EGF-LIKE 2.
 FT DOMAIN 102 139 EGF-LIKE 3.
 FT DOMAIN 140 176 EGF-LIKE 4.
 FT DOMAIN 178 216 EGF-LIKE 5.
 FT DOMAIN 218 255 EGF-LIKE 6.
 FT DOMAIN 257 293 EGF-LIKE 7.
 FT DOMAIN 295 333 EGF-LIKE 8.
 FT DOMAIN 335 371 EGF-LIKE 9.
 FT DOMAIN 372 410 EGF-LIKE 10.
 FT DOMAIN 412 450 EGF-LIKE 11.
 FT DOMAIN 452 488 EGF-LIKE 12.
 FT DOMAIN 490 526 EGF-LIKE 13.
 FT DOMAIN 528 564 EGF-LIKE 14.
 FT DOMAIN 566 601 EGF-LIKE 15.
 FT DOMAIN 603 639 EGF-LIKE 16.
 FT DOMAIN 641 676 EGF-LIKE 17.
 FT DOMAIN 678 714 EGF-LIKE 18.
 FT DOMAIN 716 751 EGF-LIKE 19.
 FT DOMAIN 753 789 EGF-LIKE 20.
 FT DOMAIN 791 827 EGF-LIKE 21.
 FT DOMAIN 829 867 EGF-LIKE 22.
 FT DOMAIN 869 905 EGF-LIKE 23.
 FT DOMAIN 907 943 EGF-LIKE 24.
 FT DOMAIN 945 981 EGF-LIKE 25.
 FT DOMAIN 983 1019 EGF-LIKE 26.
 FT DOMAIN 1021 1057 EGF-LIKE 27.
 FT DOMAIN 1059 1095 EGF-LIKE 28.
 FT DOMAIN 1097 1143 EGF-LIKE 29.
 FT DOMAIN 1145 1181 EGF-LIKE 30.
 FT DOMAIN 1183 1219 EGF-LIKE 31.
 FT DOMAIN 1221 1265 EGF-LIKE 32.
 FT DOMAIN 1267 1305 EGF-LIKE 33.
 FT DOMAIN 1307 1346 EGF-LIKE 34.
 FT DOMAIN 1348 1384 EGF-LIKE 35.
 FT DOMAIN 1387 1426 EGF-LIKE 36.
 FT DOMAIN 1449 1462 CYS-RICH.
 FT REPEAT 1445 1480 LIN/NOTCH 1.
 FT REPEAT 1481 1522 LIN/NOTCH 2.
 FT REPEAT 1523 1562 LIN/NOTCH 3.
 FT REPEAT 1917 1947 ANK 1.
 FT REPEAT 1949 1979 ANK 2.
 FT REPEAT 1983 2012 ANK 3.
 FT REPEAT 2016 2045 ANK 4.
 FT REPEAT 2049 2078 ANK 5.
 FT DISULFID 24 37 BY SIMILARITY.
 FT DISULFID 31 46 BY SIMILARITY.
 FT DISULFID 63 74 BY SIMILARITY.
 FT DISULFID 68 87 BY SIMILARITY.
 FT DISULFID 89 98 BY SIMILARITY.
 FT DISULFID 106 117 BY SIMILARITY.
 FT DISULFID 111 127 BY SIMILARITY.
 FT DISULFID 129 138 BY SIMILARITY.
 FT DISULFID 144 155 BY SIMILARITY.
 FT DISULFID 149 164 BY SIMILARITY.
 FT DISULFID 166 175 BY SIMILARITY.
 FT DISULFID 182 195 BY SIMILARITY.
 FT DISULFID 189 204 BY SIMILARITY.
 FT DISULFID 206 215 BY SIMILARITY.
 FT DISULFID 222 233 BY SIMILARITY.
 FT DISULFID 227 243 BY SIMILARITY.
 FT DISULFID 245 254 BY SIMILARITY.
 FT DISULFID 261 272 BY SIMILARITY.
 FT

FT	DISULFID	266	281	BY SIMILARITY.	DB	1740	VCGVLLSRK	1749	
FT	DISULFID	283	292	BY SIMILARITY.	RESULT	3			
FT	DISULFID	299	312	BY SIMILARITY.	NTCL_RAT		STANDARD;	PRT; 2531 AA.	
FT	DISULFID	306	321	BY SIMILARITY.	AC	Q07008;			
FT	DISULFID	323	332	BY SIMILARITY.	DT	01-NOV-1995 (Rel. 32, Created)			
FT	DISULFID	339	350	BY SIMILARITY.	DT	15-JUL-1999 (Rel. 38, Last sequence update)			
FT	DISULFID	344	359	BY SIMILARITY.	DT	01-OCT-2000 (Rel. 40, Last annotation update)			
FT	DISULFID	361	370	BY SIMILARITY.	DE	NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR.			
FT	DISULFID	376	387	BY SIMILARITY.	GN	NOTCH1.			
FT	DISULFID	381	398	BY SIMILARITY.	OS	Rattus norvegicus (Rat).			
FT	DISULFID	400	409	BY SIMILARITY.	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
FT	DISULFID	416	429	BY SIMILARITY.	OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
FT	DISULFID	423	438	BY SIMILARITY.	OC	NCBI_TaxID=10116;			
FT	DISULFID	440	449	BY SIMILARITY.	RN	[1]			
FT	DISULFID	456	467	BY SIMILARITY.	RP	SEQUENCE FROM N.A.			
FT	DISULFID	461	476	BY SIMILARITY.	RC	TISSUE=Schwann cell;			
FT	DISULFID	478	487	BY SIMILARITY.	RX	MEDLINE=92111383; PubMed=1764995;			
FT	DISULFID	494	505	BY SIMILARITY.	RA	Weinmaster G.; Roberts V.J.; Lemke G.;			
FT	DISULFID	499	514	BY SIMILARITY.	RT	"A homolog of Drosophila Notch expressed during mammalian			
FT	DISULFID	516	525	BY SIMILARITY.	RT	development.";			
FT	DISULFID	532	543	BY SIMILARITY.	RL	Development 113:199-205(1991).			
FT	DISULFID	537	552	BY SIMILARITY.	RN	[2]			
FT	DISULFID	554	563	BY SIMILARITY.	RP	REVISIONS TO 1652-1653.			
FT	DISULFID	570	580	BY SIMILARITY.	RA	Weinmaster G.;			
FT	DISULFID	575	589	BY SIMILARITY.	RL	Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.			
FT	DISULFID	591	600	BY SIMILARITY.	CC	-1- FUNCTION: REQUIRED FOR THE CORRECT DIFFERENTIATION OF A NUMBER			
FT	DISULFID	607	618	BY SIMILARITY.	CC	OF TISSUES.			
FT	DISULFID	612	627	BY SIMILARITY.	CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.			
FT	DISULFID	629	638	BY SIMILARITY.	CC	-1- DEVELOPMENTAL STAGE: IN THE EMBRYO, HIGHEST LEVELS OCCUR BETWEEN			
FT	DISULFID	645	655	BY SIMILARITY.	CC	DAYS 12 AND 14 AND DECREASE RAPIDLY TO MUCH LOWER LEVELS IN THE			
FT	DISULFID	650	664	BY SIMILARITY.	CC	ADULT.			
FT	DISULFID	666	675	BY SIMILARITY.	CC	-1- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.			
FT	DISULFID	682	693	BY SIMILARITY.	CC	-1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.			
FT	DISULFID	687	702	BY SIMILARITY.	CC	-1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.			
FT	DISULFID	704	713	BY SIMILARITY.	CC	-1- SIMILARITY: CONTAINS 5 ANK REPEATS.			
FT	DISULFID	720	730	BY SIMILARITY.	CC	-----			
FT	DISULFID	725	739	BY SIMILARITY.	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
FT	DISULFID	741	750	BY SIMILARITY.	CC	between the Swiss Institute of Bioinformatics and the EMBL outposts on			
FT	DISULFID	757	768	BY SIMILARITY.	CC	the European Bioinformatics Institute. There are no restrictions on its			
FT	DISULFID	762	777	BY SIMILARITY.	CC	use by non-profit institutions as long as its content is in no way			
FT	DISULFID	779	788	BY SIMILARITY.	CC	modified and this statement is not removed. Usage by and for commercial			
FT	DISULFID	800	815	BY SIMILARITY.	CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
FT	DISULFID	817	826	BY SIMILARITY.	CC	or send an email to license@isb-sib.ch).			
FT	DISULFID	833	844	BY SIMILARITY.	CC	-----			
FT	DISULFID	838	855	BY SIMILARITY.	CC	EMBL; X57405; CAA40667.1; -.			
FT	DISULFID	857	866	BY SIMILARITY.	DR	HSSP; P00740; 11XA.			
FT	DISULFID	873	884	BY SIMILARITY.	DR	InterPro; IPR000152; -.			
FT	DISULFID	878	893	BY SIMILARITY.	DR	InterPro; IPR000561; -.			
FT	DISULFID	895	904	BY SIMILARITY.	DR	InterPro; IPR000800; -.			
FT	DISULFID	911	922	BY SIMILARITY.	DR	InterPro; IPR001438; -.			
FT	DISULFID	916	931	BY SIMILARITY.	DR	InterPro; IPR001881; -.			
FT	DISULFID	933	942	BY SIMILARITY.	DR	InterPro; IPR002049; -.			
FT	DISULFID	987	998	BY SIMILARITY.	DR	InterPro; IPR002110; -.			
FT	DISULFID	992	1007	BY SIMILARITY.	DR	Pfam; PF00008; EGF; 36.			
FT	DISULFID	1009	1018	BY SIMILARITY.	DR	Pfam; PF00023; ank; 6.			
FT	DISULFID	1025	1036	BY SIMILARITY.	DR	Pfam; PF00066; notch; 3.			
FT	DISULFID	1030	1045	BY SIMILARITY.	DR	PRINTS; PR00010; EGFBLD.			
FT	DISULFID	1047	1056	BY SIMILARITY.	DR	PRINTS; PR00011; EGFAMIN.			
FT	DISULFID	1063	1074	BY SIMILARITY.	DR	PROSITE; PS00088; ANK_REPEAT; 4.			
FT	DISULFID	1068	1083	BY SIMILARITY.	DR	PROSITE; PS00297; ANK_REPEAT_REGION; 1.			
FT	DISULFID	1085	1094	BY SIMILARITY.	DR	PROSITE; PS00010; ASX_HYDROXYL; 22.			
FT	DISULFID	1101	1122	BY SIMILARITY.	DR	PROSITE; PS00022; EGF_1; 35.			
FT	DISULFID	1116	1131	BY SIMILARITY.	DR	PROSITE; PS01186; EGF_2; 26.			
FT	DISULFID	1133	1142	BY SIMILARITY.	DR	PROSITE; PS01187; EGF_CA; 21.			
FT	DISULFID	1149	1160	BY SIMILARITY.	DR	PROSITE; PS01186; EGF_2; 26.			
Query Match			89.18;	Score 41;	DB	1;	Length 2531;		
Best Local Similarity			90.08;	Pred. No. 4.2;					
Matches	9;	Conservative	0;	Mismatches	1;	Indels	0;	Gaps	0;
OY	1	VCGSVLLSRK	10						

FT	DOMAIN	1747	2531	CYTOPLASMIC (POTENTIAL).	FT	DISULFID	416	429	BY SIMILARITY.
FT	DOMAIN	59	58	EGF-LIKE 1.	FT	DISULFID	433	438	BY SIMILARITY.
FT	DOMAIN	20	99	EGF-LIKE 2.	FT	DISULFID	440	449	BY SIMILARITY.
FT	DOMAIN	102	139	EGF-LIKE 3.	FT	DISULFID	456	467	BY SIMILARITY.
FT	DOMAIN	140	176	EGF-LIKE 4.	FT	DISULFID	461	476	BY SIMILARITY.
FT	DOMAIN	178	216	EGF-LIKE 5.	FT	DISULFID	478	487	BY SIMILARITY.
FT	DOMAIN	218	255	EGF-LIKE 6.	FT	DISULFID	484	505	BY SIMILARITY.
FT	DOMAIN	257	293	EGF-LIKE 7.	FT	DISULFID	499	514	BY SIMILARITY.
FT	DOMAIN	295	333	EGF-LIKE 8.	FT	DISULFID	516	525	BY SIMILARITY.
FT	DOMAIN	335	371	EGF-LIKE 9.	FT	DISULFID	532	543	BY SIMILARITY.
FT	DOMAIN	372	410	EGF-LIKE 10.	FT	DISULFID	537	553	BY SIMILARITY.
FT	DOMAIN	412	450	EGF-LIKE 11.	FT	DISULFID	554	562	BY SIMILARITY.
FT	DOMAIN	452	488	EGF-LIKE 12.	FT	DISULFID	570	580	BY SIMILARITY.
FT	DOMAIN	490	526	EGF-LIKE 13.	FT	DISULFID	575	589	BY SIMILARITY.
FT	DOMAIN	528	564	EGF-LIKE 14.	FT	DISULFID	591	600	BY SIMILARITY.
FT	DOMAIN	566	601	EGF-LIKE 15.	FT	DISULFID	607	618	BY SIMILARITY.
FT	DOMAIN	603	639	EGF-LIKE 16.	FT	DISULFID	612	637	BY SIMILARITY.
FT	DOMAIN	641	676	EGF-LIKE 17.	FT	DISULFID	629	638	BY SIMILARITY.
FT	DOMAIN	678	714	EGF-LIKE 18.	FT	DISULFID	645	655	BY SIMILARITY.
FT	DOMAIN	716	751	EGF-LIKE 19.	FT	DISULFID	650	664	BY SIMILARITY.
FT	DOMAIN	753	789	EGF-LIKE 20.	FT	DISULFID	666	675	BY SIMILARITY.
FT	DOMAIN	791	827	EGF-LIKE 21.	FT	DISULFID	682	693	BY SIMILARITY.
FT	DOMAIN	829	867	EGF-LIKE 22.	FT	DISULFID	687	702	BY SIMILARITY.
FT	DOMAIN	869	905	EGF-LIKE 23.	FT	DISULFID	704	713	BY SIMILARITY.
FT	DOMAIN	907	943	EGF-LIKE 24.	FT	DISULFID	720	730	BY SIMILARITY.
FT	DOMAIN	945	981	EGF-LIKE 25.	FT	DISULFID	725	739	BY SIMILARITY.
FT	DOMAIN	983	1019	EGF-LIKE 26.	FT	DISULFID	741	750	BY SIMILARITY.
FT	DOMAIN	1021	1057	EGF-LIKE 27.	FT	DISULFID	757	767	BY SIMILARITY.
FT	DOMAIN	1059	1095	EGF-LIKE 28.	FT	DISULFID	762	778	BY SIMILARITY.
FT	DOMAIN	1097	1143	EGF-LIKE 29.	FT	DISULFID	779	788	BY SIMILARITY.
FT	DOMAIN	1145	1181	EGF-LIKE 30.	FT	DISULFID	795	806	BY SIMILARITY.
FT	DOMAIN	1183	1219	EGF-LIKE 31.	FT	DISULFID	800	815	BY SIMILARITY.
FT	DOMAIN	1221	1265	EGF-LIKE 32.	FT	DISULFID	817	826	BY SIMILARITY.
FT	DOMAIN	1267	1305	EGF-LIKE 33.	FT	DISULFID	833	844	BY SIMILARITY.
FT	DOMAIN	1307	1346	EGF-LIKE 34.	FT	DISULFID	838	855	BY SIMILARITY.
FT	DOMAIN	1348	1384	EGF-LIKE 35.	FT	DISULFID	857	866	BY SIMILARITY.
FT	DOMAIN	1387	1426	EGF-LIKE 36.	FT	DISULFID	873	884	BY SIMILARITY.
FT	DOMAIN	1449	1462	CYS-RICH.	FT	DISULFID	878	893	BY SIMILARITY.
FT	REPEAT	1917	1946	ANK 1.	FT	DISULFID	895	904	BY SIMILARITY.
FT	REPEAT	1950	1980	ANK 2.	FT	DISULFID	911	922	BY SIMILARITY.
FT	REPEAT	1984	2013	ANK 3.	FT	DISULFID	916	931	BY SIMILARITY.
FT	REPEAT	2017	2046	ANK 4.	FT	DISULFID	933	942	BY SIMILARITY.
FT	REPEAT	2050	2079	ANK 5.	FT	DISULFID	987	998	BY SIMILARITY.
FT	DISULFID	24	37	BY SIMILARITY.	FT	DISULFID	992	1007	BY SIMILARITY.
FT	DISULFID	31	46	BY SIMILARITY.	FT	DISULFID	1009	1018	BY SIMILARITY.
FT	DISULFID	48	57	BY SIMILARITY.	FT	DISULFID	1025	1036	BY SIMILARITY.
FT	DISULFID	63	74	BY SIMILARITY.	FT	DISULFID	1030	1045	BY SIMILARITY.
FT	DISULFID	68	87	BY SIMILARITY.	FT	DISULFID	1047	1056	BY SIMILARITY.
FT	DISULFID	89	98	BY SIMILARITY.	FT	DISULFID	1063	1074	BY SIMILARITY.
FT	DISULFID	106	117	BY SIMILARITY.	FT	DISULFID	1068	1083	BY SIMILARITY.
FT	DISULFID	111	127	BY SIMILARITY.	FT	DISULFID	1085	1094	BY SIMILARITY.
FT	DISULFID	129	138	BY SIMILARITY.	FT	DISULFID	1101	1122	BY SIMILARITY.
FT	DISULFID	144	155	BY SIMILARITY.	FT	DISULFID	1116	1131	BY SIMILARITY.
FT	DISULFID	149	164	BY SIMILARITY.	FT	DISULFID	1133	1142	BY SIMILARITY.
FT	DISULFID	166	175	BY SIMILARITY.	FT	DISULFID	1149	1160	BY SIMILARITY.
FT	DISULFID	182	195	BY SIMILARITY.	FT	DISULFID	1154	1169	BY SIMILARITY.
FT	DISULFID	189	204	BY SIMILARITY.	FT	DISULFID	1171	1180	BY SIMILARITY.
FT	DISULFID	206	215	BY SIMILARITY.	FT	DISULFID	1187	1198	BY SIMILARITY.
FT	DISULFID	222	233	BY SIMILARITY.	FT	DISULFID	1192	1207	BY SIMILARITY.
FT	DISULFID	227	243	BY SIMILARITY.					
FT	DISULFID	245	254	BY SIMILARITY.					
FT	DISULFID	261	272	BY SIMILARITY.					
FT	DISULFID	266	281	BY SIMILARITY.					
FT	DISULFID	283	292	BY SIMILARITY.					
FT	DISULFID	299	312	BY SIMILARITY.					
FT	DISULFID	306	321	BY SIMILARITY.					
FT	DISULFID	323	332	BY SIMILARITY.					
FT	DISULFID	339	350	BY SIMILARITY.					
FT	DISULFID	344	359	BY SIMILARITY.					
FT	DISULFID	361	370	BY SIMILARITY.					
FT	DISULFID	376	387	BY SIMILARITY.					
FT	DISULFID	381	398	BY SIMILARITY.					
FT	DISULFID	400	409	BY SIMILARITY.					
					Query Match 89.1%; Score 41; DB				
					Best Local Similarity 90.0%; Pred. No. 4.2;				
					Matches 9; Conservative 0; Mismatches 0;				
Qy	1 VGSGVLLSRK 10								
Db	1740 VGCGVLLSRK 1749								
					RESULT 4				
PHOOQ_SALTY	ID PHOOQ_SALTY				STANDARD;				
AC	P14147;				PRT; 451 AA				

Fri Sep 7 10:58:26 2001

```

DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE VIRULENCE SENSOR PROTEIN PHOQ (EC 2.7.3.-).
GN PHOQ.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 10428;
RX MEDLINE=89296942; PubMed=2544889;
RA Miller S.I., Kukral A.M., Mekalanos J.J.;
RT "A two-component regulatory system (phoP phoQ) controls Salmonella
  typhimurium virulence.";
RL INVOLVED IN THE REGULATION OF THE EXPRESSION OF GENES INVOLVED IN
  PHOQ MAY
CC -!- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM PHOQ/PHOP
  INVOLVED IN THE REGULATION OF S.TYPHIMURIUM. PHOQ MAY
CC VIRULENCE AND MACROPHAGE SURVIVAL OF S.TYPHIMURIUM. PHOQ MAY
CC FUNCTION AS A MEMBRANE-ASSOCIATED PROTEIN KINASE THAT
  PHOSPHORYLATES PHOP IN RESPONSE TO ENVIRONMENTAL SIGNALS.
CC PHOSPHORYLATES PHOP IN RESPONSE TO ENVIRONMENTAL SIGNALS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
  (PROBABLE).
CC -!- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE
  KINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M24424; AAA27189.1; -
DR PIR; B32932; VZEBPT.
DR StyGene; SGI0294; phoQ.
DR InterPro; IPR000410; -
DR InterPro; IPR000658; -
DR Pfam; PF00672; DUF5; 1.
DR Pfam; PF00512; signal; 1.
DR Sensory transduction; Transferase; Kinase; Phosphorylation;
  Transmembrane; Inner membrane; Virulence.
KW DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 17 44 POTENTIAL.
FT TRANSMEM 45 170 PERIPLASMIC (POTENTIAL).
FT DOMAIN 171 198 POTENTIAL.
FT TRANSMEM 199 451 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 236 451 TRANSMITTER DOMAIN (POTENTIAL).
FT MOD_RES 259 259 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 451 AA; 51586 MW; CE0E1E4F7BB43194 CRC64;

Query Match 82.6%; Score 38; DB 1; Length 451;
Best Local Similarity 88.9%; Pred. No. 3.1;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GSGVLLSRK 10
DB 311 GSGVLLSRE 319
|||||||

RESULT 5
LDDR_ECOLI STANDARD; PRT; 258 AA.
AC P33233;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PUTATIVE L-LACTATE DEHYDROGENASE OPERON REGULATORY PROTEIN.
GN LDDR OR LCTR.
OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=94012541; PubMed=8407843;
RA Dong J.M., Taylor J.S., Latour D.J., Iuchi S., Lin E.C.C.;
RT "Three overlapping lct genes involved in L-lactate utilization by
  Escherichia coli.";
RL J. Bacteriol. 175:6671-6678(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=94316500; PubMed=8041620;
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
  region from 76.0 to 81.5 minutes.";
RL Nucleic Acids Res. 22:2576-2586(1994).
CC -!- FUNCTION: MAY BE A REGULATORY PROTEIN FOR THE LCT GENES.
CC -!- INDUCTION: AEROBICALLY BY L-LACTATE.
CC -!- SIMILARITY: BELONGS TO THE GNTR FAMILY OF TRANSCRIPTIONAL
  REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; LI3970; AAA03584.1; -
DR EMBL; U00039; AAB18581.1; -
DR EMBL; AE000438; AAC76628.1; -
DR PIR; B49904; B49904
DR EcoGene; EGI1962; lldr.
DR InterPro; IPR000524; -
DR Pfam; PF00392; gntr; 1.
DR PRINTS; PR00035; HTHGNTR.
DR PROSITE; PS00043; HTH_GNTR_FAMILY; 1.
KW Transcription regulation; DNA-binding.
FT DNA_BIND 34 53 H-T-H MOTIF (BY SIMILARITY).
SQ SEQUENCE 258 AA; 29166 MW; 59C4643B3456079E CRC64;

Query Match 71.7%; Score 33; DB 1; Length 258;
Best Local Similarity 70.0%; Pred. No. 19;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 VSGVLLSRK 10
DB 56 VSGVLLSRR 65
|||||||

RESULT 6
FLA2_CAMJE STANDARD; PRT; 575 AA.
AC P22251;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE FLAGELLIN A.
GN FLAGA.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision;
OC Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=81116;
RX MEDLINE=91009243; PubMed=2211662;
RA Nuijten P.J., van Asten F.J., Gaastra W., van der Zeijst B.A.;

```

RT *Structural and functional analysis of two Campylobacter jejuni
RT flagellin genes.*;
RL J. Biol. Chem. 265:17798-17804(1990).
CC -1- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO
CC FORM THE FILAMENTS OF BACTERIAL FLAGELLA.
CC -1- SUBUNIT: HETEROPOLYMER OF FLAA AND FLAB.
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL FLAGELLIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J05635; AAA23024.1; -;
DR PIR; A39228; A39228;
DR InterPro; IPR001029; -;
DR Pfam; PF00700; Flagellin_C; 1.
DR PRINTS; PR00669; Flagellin_N; 1.
DR PRINTS; PR00207; FLAGELLIN.
KW Flagella.
FT INIT_MET 0 0 BY SIMILARITY.
SQ SEQUENCE 575 AA; 59398 MW; 71F192EFD0AF0DC7 CRC64;

Query Match 71.7%; Score 33; DB 1; Length 575;
Best Local Similarity 60.0%; Pred. No. 41;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 VGSGLVLSRK 10
IIII:|::|
DB 314 VGSGLANQK 323

RESULT 7
FLB2_CAMJE ID FLB2_CAMJE STANDARD; PRT; 575 AA.
AC P22252;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE FLAGELLIN B.
GN FLAB.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=81116;
RX MEDLINE=91009243; PubMed=2211662;
RA Nuijten P.J., van Asten F.J., Gastra W., van der Zeijst B.A.;
RT *Structural and functional analysis of two Campylobacter jejuni
RT flagellin genes.*;
RL J. Biol. Chem. 265:17798-17804(1990).
CC -1- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO
CC FORM THE FILAMENTS OF BACTERIAL FLAGELLA.
CC -1- SUBUNIT: HETEROPOLYMER OF FLAA AND FLAB.
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL FLAGELLIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J05635; AAA23025.1; -;
DR PIR; B39228; B39228.

DR InterPro; IPR001029; -;
DR InterPro; IPR001492; -;
DR Pfam; PF00700; Flagellin_C; 1.
DR Pfam; PF00669; Flagellin_N; 1.
DR PRINTS; PR00207; FLAGELLIN.
KW Flagella.
FT INIT_MET 0 0 BY SIMILARITY.
SQ SEQUENCE 575 AA; 59728 MW; D0531AF308A7BF1D CRC64;

Query Match 71.7%; Score 33; DB 1; Length 575;
Best Local Similarity 60.0%; Pred. No. 41;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 VGSGLVLSRK 10
IIII:|::|
DB 314 VGSGLANQK 323

RESULT 8
DCE2_HUMAN ID DCE2_HUMAN STANDARD; PRT; 585 AA.
AC Q05329;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE GLUTAMATE DECARBOXYLASE, 65 KDA ISOFORM (EC 4.1.1.15) (GAD-65)
GN (65 KDA GLUTAMIC ACID DECARBOXYLASE).
DE GAD2 OR GAD65.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92196068; PubMed=1549570;
RA Bu D.-F., Erlander M.G., Hitz B.C., Tillakaratne N.J., Kaufman D.L.,
RA Wagner-McPherson C.B., Evans G.A., Tobin A.J.;
RT "Two human glutamate decarboxylases, 65-kDa GAD and 67-kDa GAD, are
RT each encoded by a single gene."
RL Proc. Natl. Acad. Sci. U.S.A. 89:2115-2119(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94375018; PubMed=8088791;
RA Bu D.-F., Tobin A.J.;
RT "The exon-intron organization of the genes (GAD1 and GAD2) encoding
RT two human glutamate decarboxylases (GAD67 and GAD65) suggests that
RT they derive from a common ancestral GAD.";
RL Genomics 21:222-228(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreatic islets;
RX MEDLINE=92020848; PubMed=1924293;
RA Karlisen A.E., Hagopian W.A., Grubin C.E., Dube S., Distech C.M.,
RA Adler D.A., Barmer H., Mathews S., Grant F.J., Foster D.,
RA Lernmark A.;
RT "Cloning and primary structure of a human islet isoform of glutamic
RT acid decarboxylase from chromosome 10.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:8337-8341(1991).
RN [4]
RP SEQUENCE OF 6-585 FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=93185681; PubMed=7680313;
RA Mauch L., Abney C.C., Berg H., Scherbaum W.A., Liedvogel B.,
RA Northemann W.;
RT "Characterization of a linear epitope within the human pancreatic
RT 64-kDa glutamic acid decarboxylase and its autoimmune recognition by
RT sera from insulin-dependent diabetes mellitus patients.";
RL Eur. J. Biochem. 212:597-603(1993).
CC -1- FUNCTION: CATALYZES THE PRODUCTION OF GABA.
CC -1- CATALYTIC ACTIVITY: L-GLUTAMATE -> 4-AMINOBUTANOATE + CO(2).
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC

```

CC CC -1- SIMILARITY: LOCAL TO DOPA DECARBOXYLASE, AND TO PLP-TYPE HISTIDINE
CC CC DECARBOXYLASE.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; M81882; AAA62367.1; -
CC CC EMBL; M74826; AAA58491.1; -
CC CC EMBL; X69936; CAA49554.1; ALT_INIT.
CC CC EMBL; M70435; AAA52513.1; -
CC CC PIR; A41292; A41292.
CC CC PIR; PQ0158; PQ0158.
CC CC MIM; 138275; -
CC CC InterPro; IPR002129; -
CC CC Pfam; PF00282; Pyridoxal_dec; 1.
CC CC PROSITE; PS00392; DDC_GAD_HDC_YDC; 1.
CC CC Neurotransmitter biosynthesis; Lyase; Decarboxylase;
CC CC Pyridoxal phosphate; Multigene family.
CC CC BINDING 396 396 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC CC SEQUENCE 585 AA; 65411 MW; 0322509F0E2C32EE CRC64;
CC CC -----
CC CC Query Match 71.7%; Score 33; DB 1; Length 585;
CC CC Best Local Similarity 66.7%; Pred. No. 42;
CC CC Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
CC CC -----
QY 2 GSGVLLSRK 10
DB 368 GGGLLMSRK 376
| | : | | |
| | : | | |

RESULT 9
DCE2_MOUSE STANDARD; PRT; 585 AA.
AC DCE2_MOUSE STANDARD; PRT; 585 AA.
AC P48320; O35519;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE GLUTAMATE DECARBOXYLASE, 65 KDA ISOFORM (EC 4.1.1.15) (GAD-65)
DE (65 KDA GLUTAMIC ACID DECARBOXYLASE).
DE GAD2 OR GAD65.
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
ON [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=94032481; PubMed=8218409;
RA Lee D.S., Tian J., Phan T., Kaufman D.L.;
RT "Cloning and sequence analysis of a murine cDNA encoding glutamate
RT decarboxylase (GAD65).";
RL Biochim. Biophys. Acta 1216:157-160(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=97115675; PubMed=8954991;
RA Asada H., Kawamura Y., Maruyama K., Kume H., Ding R.G., Ji F.Y.,
RA Kanbara N., Kuzume H., Sanbo M., Yagi T., Obata K.;
RT "Mice lacking the 65 kDa isoform of glutamic acid decarboxylase
RT (GAD65) maintain normal levels of GAD67 and GABA in their brains but
RT are susceptible to seizures.";
RL Biochem. Biophys. Res. Commun. 229:891-895(1996).
RN [3]
RP SEQUENCE OF 175-379 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=94062679; PubMed=8243324;
RA Faulkner-Jones B.E., Cram D.S., Kun J., Harrison L.C.;

```

```

RT RT "Localization and quantitation of expression of two glutamate
RT RT decarboxylase genes in pancreatic beta-cells and other peripheral
RT RT tissues of mouse and rat."
RL Endocrinology 133:2962-2972(1993).
CC -1- FUNCTION: CATALYZES THE PRODUCTION OF GABA.
CC -1- CATALYTIC ACTIVITY: L-GLUTAMATE = 4-AMINOBUTANOATE + CO(2).
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SIMILARITY: LOCAL TO DOPA DECARBOXYLASE, AND TO PLP-TYPE HISTIDINE
CC CC DECARBOXYLASE.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; L16980; AAA93049.1; -
CC CC EMBL; DA2051; BAA22893.1; -
CC CC EMBL; S67454; -; NOT_ANNOTATED_CDS.
CC CC MGI; 95634; Gad2.
CC CC InterPro; IPR002129; -
CC CC Pfam; PF00282; Pyridoxal_dec; 1.
CC CC PROSITE; PS00392; DDC_GAD_HDC_YDC; 1.
CC CC Neurotransmitter biosynthesis; Lyase; Decarboxylase;
CC CC Pyridoxal phosphate; Multigene family.
CC CC BINDING 396 396 PYRIDOXAL PHOSPHATE (POTENTIAL).
CC CC CONFLICT 259 259 F -> S (IN REF. 2).
CC CC CONFLICT 319 319 I -> S (IN REF. 3).
CC CC CONFLICT 325 325 K -> E (IN REF. 2).
CC CC CONFLICT 499 499 P -> S (IN REF. 2).
CC CC SEQUENCE 585 AA; 65224 MW; C2F486E85123B057 CRC64;
CC CC -----
CC CC Query Match 71.7%; Score 33; DB 1; Length 585;
CC CC Best Local Similarity 66.7%; Pred. No. 42;
CC CC Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
CC CC -----
QY 2 GSGVLLSRK 10
DB 368 GGGLLMSRK 376
| | : | | |
| | : | | |

RESULT 10
DCE2_PIG STANDARD; PRT; 585 AA.
AC P48321; 1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE GLUTAMATE DECARBOXYLASE, 65 KDA ISOFORM (EC 4.1.1.15) (GAD-65)
DE (65 KDA GLUTAMIC ACID DECARBOXYLASE).
DE GAD2 OR GAD65.
GN Sus scrofa (Pig).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
ON [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=95137399; PubMed=7835711;
RA Suzuki R., Asami N., Annann E., Wagatsuma M.;
RT "Sequences of two porcine glutamic acid decarboxylases (65- and
RT 67-kDa GAD).";
RL Gene 152:257-260(1995).
CC -1- FUNCTION: CATALYZES THE PRODUCTION OF GABA.
CC -1- CATALYTIC ACTIVITY: L-GLUTAMATE = 4-AMINOBUTANOATE + CO(2).
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO GROUP II DECARBOXYLASES (DDC, GAD, HDC AND
CC CC TYRDC).

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; D31848; BAA06635.1; -
DR InterPro: IPR002129; -
DR Pfam: PF00282; pyridoxal_dec; 1.
DR PROSITE; PS00392; DDC_GAD_HDC_YDC; 1.
KW Neurotransmitter biosynthesis; Lyase; Decarboxylase;
KW Pyridoxal phosphate; Multigene family.
FT BINDING 396 396 PYRIDOXAL PHOSPHATE (POTENTIAL).
SQ SEQUENCE 585 AA; 65388 MW; 4FF2810637671B6B CRC64;

Query Match 71.7%; Score 33; DB 1; Length 585;
Best Local Similarity 66.7%; Pred. No. 42;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSGVLLSRK 10
| : : : :
DB 368 GGGLLSRK 376

RESULT 11
DCE2_RAT STANDARD; PRT; 585 AA.
ID DCE2_RAT
AC Q05683;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE GLUTAMATE DECARBOXYLASE, 65 KDA ISOFORM (EC 4.1.1.15) (GAD-65)
DE (65 KDA GLUTAMIC ACID DECARBOXYLASE).
GN GAD2 OR GAD65.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Hippocampus;
RX MEDLINE=91299343; PubMed=2069816;
RA Erlanger M.G., Tillakaratne N.J., Feldblum S., Patel N.,
RA Tobin A.J.;
RT "Two genes encode distinct glutamate decarboxylases.";
RL Neuron 7:91-100(1991).
CC -!- FUNCTION: CATALYZES THE PRODUCTION OF GABA.
CC -!- CATALYTIC ACTIVITY: L-GLUTAMATE -> 4-AMINOBUTANOATE + CO(2).
CC -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -!- SIMILARITY: LOCAL TO DOPA DECARBOXYLASE, AND TO PLP-TYPE HISTIDINE
CC DECARBOXYLASE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M72422; AAA63488.1; -
DR InterPro: IPR002129; -
DR Pfam: PF00282; pyridoxal_dec; 1.
DR PROSITE; PS00392; DDC_GAD_HDC_YDC; 1.
KW Neurotransmitter biosynthesis; Lyase; Decarboxylase;
KW Pyridoxal phosphate; Multigene family.
FT BINDING 396 396 PYRIDOXAL PHOSPHATE (POTENTIAL).
SQ SEQUENCE 585 AA; 65402 MW; C04040B78A7B37D1 CRC64;

Query Match 71.7%; Score 33; DB 1; Length 585;
Best Local Similarity 66.7%; Pred. No. 42;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSGVLLSRK 10
| : : : :
DB 368 GGGLLSRK 376

RESULT 12
DCE1_MOUSE STANDARD; PRT; 593 AA.
ID DCE1_MOUSE
AC P48318; O08685;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE GLUTAMATE DECARBOXYLASE, 67 KDA ISOFORM (EC 4.1.1.15) (GAD-67)
DE (67 KDA GLUTAMIC ACID DECARBOXYLASE).
GN GAD1 OR GAD67.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Katarova Z., Szabo G., Mugnaini E., Greenspan R.;
RT "Molecular identification of the 62 kd form of glutamic acid
RT decarboxylase from the mouse.";
RL Eur. J. Neurosci. 2:190-202(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RA Aust G., Steinbrenner H., Thamm B., Rost A.K., Seissler J.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 198-403 FROM N.A.
RX MEDLINE=94062679; PubMed=8243324;
RA Faulkner-Jones B.E., Gram D.S., Kun J., Harrison L.C.;
RT "Localization and quantitation of expression of two glutamate
RT decarboxylase genes in pancreatic beta-cells and other peripheral
RT tissues of mouse and rat.";
RL Endocrinology 133:2962-2972(1993).
CC -!- FUNCTION: CATALYZES THE PRODUCTION OF GABA.
CC -!- CATALYTIC ACTIVITY: L-GLUTAMATE -> 4-AMINOBUTANOATE + CO(2).
CC -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO GROUP II DECARBOXYLASES (DDC, GAD, HDC AND
CC TYRDC).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; Z49976; CAA90277.1; -
DR EMBL; Y12257; CAA72934.1; -
DR EMBL; S67453; -; NOT_ANNOTATED_CDS.
DR MGD; MGI:95632; Gad1.
DR InterPro: IPR002129; -
DR Pfam: PF00282; pyridoxal_dec; 1.
DR PROSITE; PS00392; DDC_GAD_HDC_YDC; 1.
KW Neurotransmitter biosynthesis; Lyase; Decarboxylase;
KW Pyridoxal phosphate; Multigene family.
FT BINDING 404 404 PYRIDOXAL PHOSPHATE (POTENTIAL).
FT CONFLICT 133 133 H -> R (IN REF. 2).
FT CONFLICT 234 234 E -> K (IN REF. 3).
FT

FT CONFLICT 258 258 S -> T (IN REF. 1).
FT CONFLICT 360 360 D -> S (IN REF. 1).
FT CONFLICT 461 461 N -> F (IN REF. 2).
FT CONFLICT 554 554 A -> G (IN REF. 2).
FT CONFLICT 575 575 S -> T (IN REF. 2).
FT CONFLICT 583 583 T -> I (IN REF. 2).
SQ SEQUENCE 593 AA; 66584 MW; C257A0BEF63FCDDC CRC64;

Query Match 71.7%; Score 33; DB 1; Length 593;
Best Local Similarity 66.7%; Pred. No. 42;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSGVLLSRK 10
| : : : :
DB 376 GGLLMSRK 384

RESULT 13
ID DCEL_RAT STANDARD; PRT; 593 AA.
AC P18088;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE GLUTAMATE DECARBOXYLASE, 67 KDA ISOFORM (EC 4.1.1.15) (GAD-67)
DE (67 KDA GLUTAMIC ACID DECARBOXYLASE).
GN GAD1 OR GAD67.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP MEDLINE=91014554; PubMed=2170798;
RA Wyborski R.J., Bond R.W., Gottlieb D.I.;
RT "Characterization of a cDNA coding for rat glutamic acid
decarboxylase";
RL Brain Res. Mol. Brain Res. 8:193-198(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90132703; PubMed=2299361;
RA Julien J.F., Samama P., Mallet J.;
RT "rat brain glutamic acid decarboxylase sequence deduced from a cloned
cDNA";
RL J. Neurochem. 54:703-705(1990).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=92020930; PubMed=1924335;
RA Michelsen B.K., Petersen J.S., Boel E., Moldrup A., Dyrberg T.,
Madsen O.D.;
RT "Cloning, characterization, and autoimmune recognition of rat islet
glutamic acid decarboxylase in insulin-dependent diabetes mellitus";
RL Proc. Natl. Acad. Sci. U.S.A. 88:8754-8758(1991).
CC -1- FUNCTION: CATALYZES THE PRODUCTION OF GABA.
CC -1- CATALYTIC ACTIVITY: L-GLUTAMATE = 4-AMINOBUTANOATE + CO(2).
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SIMILARITY: BELONGS TO GROUP II DECARBOXYLASES (DDC, GAD, HDC AND
TYRDC).

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M34445; AAC42037.1;
DR EMBL; X57572; CAA40800.1;
DR EMBL; X57573; CAA40801.1;
DR EMBL; M76177; AAA41184.1;

DR PIR; A41367; A41367.
DR PIR; A43756; A43756.
DR InterPro; IPR002129;
DR Pfam; PF00282; Pyridoxal_deC; 1.
DR PROSITE; PS00392; DDC_GAD_HDC_YDC; 1. Decarboxylase;
KW Neurotransmitter biosynthesis; Lyase;
KW Pyridoxal phosphate; Multigene family.
FT BINDING 404 404 PYRIDOXAL PHOSPHATE (POTENTIAL).
FT CONFLICT 103 103 L -> V (IN REF. 2).
FT CONFLICT 284 284 F -> S (IN REF. 2).
FT CONFLICT 287 288 EH -> AD (IN REF. 2).
FT CONFLICT 344 345 AG -> EA (IN REF. 2).
FT CONFLICT 347 347 T -> I (IN REF. 2).
FT CONFLICT 352 353 FD -> LE (IN REF. 2).
FT CONFLICT 380 380 L -> R (IN REF. 2).
SQ SEQUENCE 593 AA; 66640 MW; EF83239C30301F69 CRC64;

Query Match 71.7%; Score 33; DB 1; Length 593;
Best Local Similarity 66.7%; Pred. No. 42;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSGVLLSRK 10
| : : : :
DB 376 GGLLMSRK 384

RESULT 14
ID DCEL_FELCA STANDARD; PRT; 594 AA.
AC P14748;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE GLUTAMATE DECARBOXYLASE, 67 KDA ISOFORM (EC 4.1.1.15) (GAD-67)
DE (67 KDA GLUTAMIC ACID DECARBOXYLASE).
GN GAD1 OR GAD67.
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Occipital cortex;
MEDLINE=87310623; PubMed=3453123;
RA Kobayashi Y., Kaufman D.L., Tobin A.J.;
RT "Glutamic acid decarboxylase cDNA: nucleotide sequence encoding an
enzymatically active fusion protein";
RL J. Neurosci. 7:2768-2772(1987).
CC -1- FUNCTION: CATALYZES THE PRODUCTION OF GABA.
CC -1- CATALYTIC ACTIVITY: L-GLUTAMATE = 4-AMINOBUTANOATE + CO(2).
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SIMILARITY: BELONGS TO GROUP II DECARBOXYLASES (DDC, GAD, HDC AND
TYRDC).

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M18629; AAA51430.1;
DR PIR; A45671; A45671.
DR PIR; A46758; A46758.
DR InterPro; IPR002129;
DR Pfam; PF00282; Pyridoxal_deC; 1.
DR PROSITE; PS00392; DDC_GAD_HDC_YDC; 1. Decarboxylase;
KW Neurotransmitter biosynthesis; Lyase;
KW Pyridoxal phosphate; Multigene family.
FT BINDING 405 405 PYRIDOXAL PHOSPHATE (POTENTIAL).

SO SEQUENCE 594 AA; 66824 MW; EE1C8D928BC0BD02 CRC64;

Query Match 71.7%; Score 33; DB 1; Length 594;
Best Local Similarity 66.7%; Pred. No. 42;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 2 GSGVLLSRK 10
| :|:|:|
Db 377 GGLLSMRK 385

RESULT 15
DCE1_HUMAN STANDARD; PRT; 594 AA.
AC Q9259;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE GLUTAMATE DECARBOXYLASE, 67 KDA ISOFORM (EC 4.1.1.15) (GAD-67)
DE (67 KDA GLUTAMIC ACID DECARBOXYLASE).
GN GAD1 OR GAD.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=92196068; PubMed=1549570;
RA Bu D.-F., Erlander M.G., Hitz B.C., Tillakaratne N.J., Kaufman D.L.,
Wagner-McPherson C.B., Evans G.A., Tobin A.J.;
RT "Two human glutamate decarboxylases, 65-kDa GAD and 67-kDa GAD, are
each encoded by a single gene."
RL Proc. Natl. Acad. Sci. U.S.A. 89:2115-2119(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94375018; PubMed=8088791;
RA Bu D.-F., Tobin A.J.;
RT "The exon-intron organization of the genes (GAD1 and GAD2) encoding
two human glutamate decarboxylases (GAD67 and GAD65) suggests that
they derive from a common ancestral GAD."
RL Genomics 21:222-228(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=93080286; PubMed=1339255;
RA Kelly C.D., Edwards Y., Johnstone A.P., Harfst E., Nogradi A.,
Nussey S.S., Povey S., Carter N.D.;
RT "Nucleotide sequence and chromosomal assignment of a cDNA encoding
the large isoform of human glutamate decarboxylase."
RL Ann. Hum. Genet. 56:255-265(1992).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=93282844; PubMed=8507202;
RA Yamashita K., Cram D.S., Harrison L.C.;
RT "Molecular cloning of full-length glutamic acid decarboxylase 67 from
human pancreas and islets."
RL Biochem. Biophys. Res. Commun. 192:1347-1352(1993).
RN [5]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreatic islets;
RX MEDLINE=93282845; PubMed=8507203;
RA Kawasaki K., Moriuchi R., Watanabe M., Saitoh K., Brunicardi F.C.,
Watt P.C., Yamaguchi T., Mullen Y., Akazawa S., Miyamoto T.;
RT "Cloning and expression of large isoform of glutamic acid
decarboxylase from human pancreatic islet."
RL Biochem. Biophys. Res. Commun. 192:1353-1359(1993).
RN [6]
RP SEQUENCE FROM N.A.
RA Giorda R., Peakman M., Vergani D., Trucco M.;
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
RN [7]

RP SEQUENCE OF 218-397 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=91248209; PubMed=2039509;
RA Cram D.S., Barnett L.D., Joseph J.L., Harrison L.C.;
RT "Cloning and partial nucleotide sequence of human glutamic acid
decarboxylase cDNA from brain and pancreatic islets."
RL Biochem. Biophys. Res. Commun. 176:1239-1244(1991).
RN [8]
RP SEQUENCE OF 527-594 FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=90355986; PubMed=1697032;
RA Persson H., Pelto-Huikko M., Metsis M., Soeder O., Brene S.,
Skog S., Hoekfelt T., Ritzén E.M.;
RT "Expression of the neurotransmitter-synthesizing enzyme glutamic acid
decarboxylase in male germ cells."
RL Mol. Cell. Biol. 10:4701-4711(1990).
CC 1- FUNCTION: CATALYZES THE PRODUCTION OF GABA.
CC 1- CATALYTIC ACTIVITY: L-GLUTAMATE -> 4-AMINOBUTANOATE + CO(2).
CC 1- COFACTOR: PYRIDOXAL PHOSPHATE.
CC 1- SUBUNIT: HOMODIMER.
CC 1- SIMILARITY: BELONGS TO GROUP II DECARBOXYLASES (DDC, GAD, HDC AND
TYRDC).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; M81883; AAA62368.1; -
DR EMBL; L16888; AAB59427.1; -
DR EMBL; 222750; CAA80435.1; -
DR EMBL; S61897; AAB26937.1; -
DR EMBL; S61898; AAB26938.1; -
DR EMBL; M86522; AAA35900.1; -
DR EMBL; M70434; AAA52512.1; -
DR EMBL; M55574; AAA72938.1; -
DR EMBL; A28074; CAA01913.1; -
DR PIR; PQ0157; PQ0157.
DR MIM; 605363; -
DR InterPro; IPR002129; -
DR Pfam; PF00282; pyridoxal_dec; 1.
DR PROSITE; PS00392; DDC_GAD_HDC_YDC; 1.
KW Neurotransmitter biosynthesis; Lyase; Decarboxylase;
KW Pyridoxal phosphate; Multigene family.
FT BINDING 405 405 PYRIDOXAL PHOSPHATE (POTENTIAL).
FT CONFLICT 9 9 MISSING (IN REF. 6).
FT CONFLICT 16 17 GA -> EP (IN REF. 3).
FT CONFLICT 17 17 A -> Q (IN REF. 6).
FT CONFLICT 18 18 D -> N (IN REF. 5).
FT CONFLICT 31 31 T -> N (IN REF. 5).
FT CONFLICT 68 68 K -> R (IN REF. 1 AND 2).
FT CONFLICT 116 116 F -> L (IN REF. 5).
FT CONFLICT 136 136 T -> A (IN REF. 6).
FT CONFLICT 140 140 D -> E (IN REF. 6).
FT CONFLICT 142 142 H -> R (IN REF. 6).
FT CONFLICT 155 155 N -> T (IN REF. 5).
FT CONFLICT 206 206 T -> N (IN REF. 3).
FT CONFLICT 302 302 G -> C (IN REF. 5).
FT CONFLICT 436 436 F -> L (IN REF. 4).
FT CONFLICT 512 512 N -> S (IN REF. 4).
FT CONFLICT 477 477 E -> G (IN REF. 5).
FT CONFLICT 492 492 A -> G (IN REF. 5).
FT CONFLICT 565 565 F -> L (IN REF. 3).
SQ SEQUENCE 594 AA; 66896 MW; 6D761C471C81FDAE CRC64;

Query Match 71.7%; Score 33; DB 1; Length 594;
Best Local Similarity 66.7%; Pred. No. 42;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GSGVLSRK 10
| | : | | |
Db 377 GGGLMSRK 385

Search completed: September 6, 2001, 16:51:13
Job time: 817 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:45:56 ; Search time 134.15 Seconds
(without alignments)
5.678 Million cell updates/sec

Title: US-09-603-713-30

Perfect score: 46

Sequence: 1 VSGGVLLSRK 10

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_68:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	89.1	2531	2 S18188	notch protein homo
2	41	89.1	2531	2 A46019	Notch-1 protein -
3	41	89.1	2555	2 A40043	notch protein homo
4	38	82.6	487	1 V2EBPT	sensor kinase phoQ
5	35	76.1	386	2 G86652	GTP-binding protei
6	34	73.9	80	2 T29969	hypothetical prote
7	34	73.9	204	2 F84184	hypothetical prote
8	34	73.9	366	2 H75180	hypothetical prote
9	33	71.7	82	2 D72280	conserved hypothet
10	33	71.7	206	2 I53274	glutamate decarbox
11	33	71.7	227	2 D84058	hypothetical prote
12	33	71.7	258	1 B49904	probable regulator
13	33	71.7	258	2 C86036	transcription regu
14	33	71.7	314	2 G71367	conserved hypothet
15	33	71.7	352	2 E82307	proteinase Dags VC
16	33	71.7	576	2 A39228	flagellin A - Camp
17	33	71.7	576	2 B39228	flagellin B - Camp
18	33	71.7	585	1 S38533	glutamate decarbox
19	33	71.7	585	1 A41292	glutamate decarbox
20	33	71.7	585	1 JH0423	glutamate decarbox
21	33	71.7	585	2 S61534	glutamate decarbox
22	33	71.7	585	2 JC4064	glutamate decarbox
23	33	71.7	593	1 A41367	glutamate decarbox
24	33	71.7	593	2 S51776	glutamate decarbox
25	33	71.7	593	2 S48135	glutamate decarbox
26	33	71.7	594	1 B41935	glutamate decarbox
27	33	71.7	594	1 A46758	glutamate decarbox
28	33	71.7	594	2 S51775	glutamate decarbox
29	33	71.7	594	2 JC4065	glutamate decarbox

30 33 71.7 1350 2 T42697 hypothetical prote
31 33 71.7 2515 2 S47008 tenascin-like prot
32 32 69.6 147 2 H70630 hypothetical prote
33 32 69.6 462 2 F83510 flagellar hook pro
34 32 69.6 463 2 F70627 probable naru prot
35 32 69.6 548 2 T39968 probable 1-pyrroli
36 32 69.6 660 2 S73673 hypothetical prote
37 32 69.6 714 2 T35770 hypothetical prote
38 32 69.6 1577 2 T15851 hypothetical prote
39 31 67.4 117 1 BVECAR arsenical resistan
40 31 67.4 137 2 F69507 c-myc binding prot
41 31 67.4 234 2 S61309 nitrate reductase
42 31 67.4 296 2 T38492 hypothetical prote
43 31 67.4 302 2 E70654 hypothetical prote
44 31 67.4 313 2 T43044 hypothetical prote
45 31 67.4 333 2 T06865 hypothetical prote

ALIGNMENTS

RESULT 1
S18188
notch protein hcmolog - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 13-Aug-1999
C:Accession: S18188
R:Weinmaster, G.; Roberts, V.J.; Lemke, G.
Development 113, 199-205, 1991
A:Title: A homolog of Drosophila Notch expressed during mammalian development.
A:Reference number: S18188; MUID:92111383
A:Accession: S18188
A:Molecule type: mRNA
A:Residues: 1-2531 <WEI>
A:Cross-references: EMBL:X57405; NID:g57634; PID:g57635
C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homol
F:987-1018/Domain: EGF homology <EGF1>
F:1025-1056/Domain: EGF homology <EGF>
F:1233-1264/Domain: EGF homology <EGF2>
F:1917-1949/Domain: ankyrin repeat homology <AN1>
F:1950-1982/Domain: ankyrin repeat homology <AN2>
F:1984-2016/Domain: ankyrin repeat homology <AN3>
F:2017-2049/Domain: ankyrin repeat homology <AN4>
F:2050-2082/Domain: ankyrin repeat homology <AN5>

Query Match 89.1%; Score 41; DB 2; Length 2531;
Best Local Similarity 90.0%; Pred.No. 7.5;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VSGGVLLSRK 10
|||
DB 1740 VCGVLLSRK 1749

RESULT 2
A46019
Notch-1 protein - mouse
N:Alternate names: notch protein
C:Species: Mus musculus (house mouse)
C:Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 20-Sep-1999
C:Accession: A46019; S25144
R:del Amo, F.F.; Gendron-Maguire, M.; Swiatek, P.J.; Jenkins, N.A.; Copeland, N.G.; G
Genomics 15, 255-264, 1993
A:Title: Cloning, analysis, and chromosomal localization of Notch-1, a mouse homolog
A:Reference number: A46019; MUID:93194170
A:Accession: A46019
A:Status: not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-2531
A:Cross-references: GB:211886; GB:S47228; NID:g288502; PIDN:CAA77941.1; PID:g288503
A>Note: sequence extracted from NCBI backbone (NCBIP:127318)
R:Franco del Amo, F.; Smith, D.E.; Swiatek, P.J.; Gendron-Maguire, M.; Greenspan, R.J

hypothetical protein PAB0416 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
A:Accession: H75130
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome 8
A:Reference number: A75001
A:Accession: H75130
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-366 <KAW>
A:Cross-references: GB:AJ248284; GB:AL096836; NID:g5457730; PIDN:CAB49527.1; PID:g545
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB0416
C:Superfamily: Pyrococcus abyssi hypothetical protein PAB0416

Query Match 73.9%; Score 34; DB 2; Length 366;
Best Local Similarity 66.7%; Pred. No. 29;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 GSGVLLSRK 10
||:||||:
Db 237 GNGVLITRK 245

RESULT 9
D72280
conserved hypothetical protein - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
A:Accession: D72280
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
A:Reference number: A72200; MUID:99287316
A:Accession: D72280
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-82 <ARN>
A:Cross-references: GB:AE001778; GB:AE000512; NID:g4981757; PIDN:AAD36281.1; PID:g498
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM1206

Query Match 71.7%; Score 33; DB 2; Length 82;
Best Local Similarity 87.5%; Pred. No. 10;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VCSGVLLS 8
||:||||:
Db 9 VGAGVLLS 16

RESULT 10
I53274
glutamate decarboxylase, 67K, brain - mouse (fragment)
C:Species: Mus sp. (mouse)
C:Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Feb-1997
A:Accession: I53274
R:Faulkner-Jones, B.E.; Cram, D.S.; Kun, J.; Harrison, L.C.
Endocrinology 133, 2962-2972, 1993
A:Title: Localization and quantitation of expression of two glutamate decarboxylase 9
A:Reference number: I53274; MUID:94062679
A:Accession: I53274
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-206 <RES>

Query Match 76.1%; Score 35; DB 2; Length 386;
Best Local Similarity 60.0%; Pred. No. 19;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 VCSGVLLSRK 10
||:||||:
Db 111 VCGGIMLSRQ 120

RESULT 6
T29969
hypothetical protein ZK682.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
A:Accession: T29969
R:Du, Z.; Le, T.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid ZK682.
A:Reference number: Z20714
A:Accession: T29969
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-80 <DUZ>
A:Cross-references: EMBL:U41110; PIDN:AAA82416.1; CESP:ZK682.3
C:Genetics:
A:Gene: CESP:ZK682.3
A:Introns: 64/2

Query Match 73.9%; Score 34; DB 2; Length 80;
Best Local Similarity 70.0%; Pred. No. 6.1;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VCSGVLLSRK 10
||:||||:
Db 22 VCGVLLLR 31

RESULT 7
F84184
hypothetical protein Vng0244h [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
A:Accession: F84184
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483
A:Accession: F84184
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-204 <STO>
A:Cross-references: GB:AE004437; NID:g10579884; PIDN:AAG18842.1; GSPDB:GN00138
C:Genetics:
A:Gene: VNG0244H

Query Match 73.9%; Score 34; DB 2; Length 204;
Best Local Similarity 70.0%; Pred. No. 16;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 VCSGVLLSRK 10
||:||||:
Db 40 ICGVLLSNK 49

RESULT 8
H75180

A:Cross-references: GB:S67453; NID:q456853
C:Superfamily: human glutamate decarboxylase

Query Match 71.7%; Score 33; DB 2; Length 206;
Best Local Similarity 66.7%; Pred. No. 26;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSGVLLSRK 10
| | | | |
Db 179 GGGVLLSRK 187

RESULT 11
DB4058
hypothetical protein BH3268 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
C:Accession: DB4058
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: AB3650; MUID:20263314
A:Accession: DB4058
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-227 <STO>
A:Cross-references: GB:AP001518; GB:BA000004; NID:gi10175792; PIDN:BA06987.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH3268
C:Superfamily: yohK protein

Query Match 71.7%; Score 33; DB 2; Length 227;
Best Local Similarity 66.7%; Pred. No. 29;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGSGVLLSR 9
| | | | |
Db 102 VGSGVLLMAK 110

RESULT 12
B49904
probable regulatory protein lctR - Escherichia coli
C:Species: Escherichia coli
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: B49904; S47825; F65160
R:Dong, J.M.; Taylor, J.S.; Latour, D.J.; Iuchi, S.; Lin, E.C.C.
J. Bacteriol. 175, 6671-6678, 1993
A:Title: Three overlapping lct genes involved in L-lactate utilization by Escherichia co
A:Reference number: A49904; MUID:94012541
A:Accession: B49904
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-258 <DON>
A:Cross-references: GB:L13970; NID:g404692; PIDN:AAA03584.1; PID:g404694
R:Plunkett, G.
submitted to the EMBL Data Library, March 1994
A:Reference number: S47666
A:Accession: S47825
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-258 <PU>
A:Cross-references: EMBL:U00039; NID:g456582; PID:g456742
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: F65160
A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA
A:Residues: 1-258 <BLAT>
A:Cross-references: GB:AE000438; GB:U00096; NID:q2367251; PIDN:AAC76628.1; PID:g17900
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: lctR
C:Superfamily: regulatory protein fadR
C:Keywords: DNA binding; transcription regulation

Query Match 71.7%; Score 33; DB 1; Length 258;
Best Local Similarity 70.0%; Pred. No. 33;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 VSGVLLSRK 10
| | | | |
Db 56 VSEGVLSSRR 65

RESULT 13
C86036
transcription regulator [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: C86036
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: C86036
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-258 <STO>
A:Cross-references: GB:AE005174; NID:gi2518354; PIDN:AA058751.1; GSPDB:GN00145; UWGP:
A:Experimental source: strain O157:H7, substrain EDL333
C:Genetics:
A:Gene: lldR
C:Superfamily: regulatory protein fadR

Query Match 71.7%; Score 33; DB 2; Length 258;
Best Local Similarity 70.0%; Pred. No. 33;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 VGSGVLLSRK 10
| | | | |
Db 56 VSEGVLSSRR 65

RESULT 14
G71367
conserved hypothetical protein TP0086 - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C:Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 05-Nov-1999
C:Accession: G71367
R:Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; G
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; M
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A:Reference number: A71250; MUID:98332770
A:Accession: G71367
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-314 <COL>
A:Cross-references: GB:AE001193; GB:AE000520; NID:g3322343; PIDN:AAC65081.1; PID:g332
A:Experimental source: strain Nichols
C:Genetics:
A:Gene: TP0086

Query Match 71.7%; Score 33; DB 2; Length 314;

Best Local Similarity 70.0%; Pred. No. 41;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VCGSVLLSRK 10
II: I:IIII
DB 69 VCAKVILSRK 78

RESULT 15
E82307
proteinase Degs VC0565 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: E82307
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Olin, H.; Dragoi, I.; Sellers, H.
I, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833
A:Accession: E82307
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-352 <HEI>
A:Cross-references: GB:AE004142; GB:AE003852; NID:g9654988; PIDN:AAF93733.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC0565
A:Map position: 1
C:Superfamily: Escherichia coli trypsin-like proteinase

Query Match 71.7%; Score 33; DB 2; Length 352;
Best Local Similarity 60.0%; Pred. No. 46;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 VCGSVLLSRK 10
:IIII:II
DB 79 LCGGVIVSEK 88

Search completed: September 6, 2001, 16:45:56
Job time: 500 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:39:40 ; Search time 113.12 seconds
(without alignments)
1.820 Million cell updates/sec

Title: US-09-603-713-30
Perfect score: 46
Sequence: 1 VSGVLLSRK 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PTCUS_COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	89.1	1068	1	US-08-537-210A-2
2	41	89.1	1068	4	US-09-113-825-2
3	41	89.1	1078	1	US-08-264-534-32
4	41	89.1	1078	1	US-08-083-590A-11
5	41	89.1	1078	1	US-08-465-500-32
6	41	89.1	1078	2	US-08-346-128-32
7	41	89.1	1078	3	US-08-532-384-11
8	41	89.1	1078	3	US-08-893-828-32
9	41	89.1	2556	1	US-08-185-432-17
10	41	89.1	2556	1	US-08-083-590A-20
11	41	89.1	2556	3	US-08-532-384-20
12	33	71.7	9	3	US-08-159-339A-979
13	33	71.7	10	3	US-08-159-339A-994
14	33	71.7	166	2	US-08-592-696-5
15	33	71.7	166	2	US-09-027-536-5
16	33	71.7	166	3	US-09-028-148-5
17	33	71.7	181	2	US-08-308-952-3
18	33	71.7	181	2	US-08-308-952-4
19	33	71.7	181	2	US-08-308-952-5
20	33	71.7	181	4	US-09-124-141-3
21	33	71.7	181	4	US-09-124-141-4
22	33	71.7	181	4	US-09-124-141-5
23	33	71.7	197	4	US-09-124-141-19
24	33	71.7	206	4	US-09-124-141-11
25	33	71.7	371	2	US-08-829-026A-6
26	33	71.7	584	6	5475086-4
27	33	71.7	584	1	US-08-161-290-1
28	33	71.7	584	1	US-08-161-290-2
29	33	71.7	584	2	US-08-450-755-1
30	33	71.7	584	2	US-08-450-755-2
31	33	71.7	585	1	US-08-117-907-2
32	33	71.7	585	1	US-08-485-718-11
33	33	71.7	585	1	US-08-485-718-12
34	33	71.7	585	1	US-08-485-718-13
35	33	71.7	585	2	US-08-484-530-57
36	33	71.7	585	2	US-08-484-530-58
37	33	71.7	585	2	US-08-484-530-59
38	33	71.7	585	2	US-08-494-824-2
39	33	71.7	585	2	US-08-827-618A-57
40	33	71.7	585	2	US-08-827-618A-58
41	33	71.7	585	2	US-08-827-618A-59
42	33	71.7	585	3	US-08-483-952A-57
43	33	71.7	585	3	US-08-483-952A-58
44	33	71.7	585	3	US-08-483-952A-59
45	33	71.7	585	6	5475086-7

ALIGNMENTS

RESULT 1

US-08-537-210A-2
; Sequence 2, Application US/08537210A
; Patent No. 5780300

GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon
; APPLICANT: Fortini, Mark
; APPLICANT: Matsuno, Kenji
; TITLE OF INVENTION: MANIPULATION OF NON-TERMINALLY
; TITLE OF INVENTION: DIFFERENTIATED CELLS USING THE NOTCH PATHWAY
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036/2711

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/537,210A
; FILING DATE: 29-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-027
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864
; TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1068 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Human N1 (TAN-1)
; LOCATION: 1152...2219
; OTHER INFORMATION: highly conserved ankyrin repeat
; OTHER INFORMATION: region of No. 5780300ch

US-08-537-210A-2

Query Match 89.1%; Score 41; DB 1; Length 1068;
Best Local Similarity 90.0%; Pred. No. 5.4;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VCGSVLLSRK 10
|| |||||
Db 600 VCGSVLLSRK 609

RESULT 2

US-09-113-825-2
Sequence 2, Application US/09113825
Patent No. 6149902
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, Spyridon
APPLICANT: Fortini, Mark
APPLICANT: Matsuno, Kenji
TITLE OF INVENTION: MANIPULATION OF NON-TERMINALLY
TITLE OF INVENTION: DIFFERENTIATED CELLS USING THE NOTCH PATHWAY
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036/2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/113,825
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/537,210
FILING DATE: 29-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: Mirock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-027
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1068 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Human.N1 (TAN-1)
LOCATION: 1152...2219
OTHER INFORMATION: Highly conserved ankyrin repeat
OTHER INFORMATION: region of No. 6149902ch

US-09-113-825-2

Query Match 89.1%; Score 41; DB 4; Length 1068;
Best Local Similarity 90.0%; Pred. No. 5.4;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VCGSVLLSRK 10
|| |||||
Db 600 VCGSVLLSRK 609

RESULT 3

US-08-264-534-32

Sequence 32, Application US/08264534
Patent No. 5648464
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, Spyridon et al.

TITLE OF INVENTION: Human No. 5648464ch And Delta, Binding Domains
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/264,534
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/695,189
FILING DATE: 03-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Mirock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-004
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 8698864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 1078 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-264-534-32

Query Match 89.1%; Score 41; DB 1; Length 1078;
Best Local Similarity 90.0%; Pred. No. 5.5;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VCGSVLLSRK 10
|| |||||
Db 297 VCGSVLLSRK 306

RESULT 4

US-08-083-590A-11
Sequence 11, Application US/08083590A
Patent No. 5786158
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, S. et al.
TITLE OF INVENTION: Therapeutic And Diagnostic Methods
TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And
TITLE OF INVENTION: Nucleic Acids
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

Query Match 89.1%; Score 41; DB 4; Length 1068;
Best Local Similarity 90.0%; Pred. No. 5.4;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VCGSVLLSRK 10
|| |||||
Db 600 VCGSVLLSRK 609

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/083,590A
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 8698864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1078 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-083-590A-11

Query Match 89.1%; Score 41; DB 1; Length 1078;
Best Local Similarity 90.0%; Pred. No. 5.5;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VGSGLVLSRK 10
|| |||||
DB 297 VGGVLLSRK 306

RESULT 5
US-08-465-500-32
; Sequence 32, Application US/08465500
; Patent No. 5789195
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon
; APPLICANT: Muskavitch, Marc A.T.
; APPLICANT: Fehon, Richard G.
; APPLICANT: Rebay, Ilaria
; APPLICANT: Blaumueller, Cristine M.
; APPLICANT: Shepard, Scott B.
; TITLE OF INVENTION: HUMAN NOTCH AND DELTA, BINDING DOMAINS
; TITLE OF INVENTION: IN TOPOTHYMIC PROTEINS, AND METHODS BASED THEREON
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,500
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-034
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1078 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-465-500-32

Query Match 89.1%; Score 41; DB 1; Length 1078;
Best Local Similarity 90.0%; Pred. No. 5.5;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VGSGLVLSRK 10
|| |||||
DB 297 VGGVLLSRK 306

RESULT 6
US-08-346-128-32
; Sequence 32, Application US/08346128
; Patent No. 5856441
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon et al.
; TITLE OF INVENTION: Human No. 5856441ch And Delta, Binding Domains
; TITLE OF INVENTION: In Topothymic Proteins, And Methods Based Thereon
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/346,128
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/879,038
; FILING DATE: 30-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-009
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 8698864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1078 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-346-128-32

Query Match 89.1%; Score 41; DB 2; Length 1078;
Best Local Similarity 90.0%; Pred. No. 5.5;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VGSGLVLSRK 10
|| |||||
DB 297 VGGVLLSRK 306

RESULT 7
US-08-532-384-11
; Sequence 11, Application US/08532384
; Patent No. 6083904
; GENERAL INFORMATION:

APPLICANT: Artavanis-Tsakonas, S. et al.
TITLE OF INVENTION: Therapeutic And Diagnostic Methods
TITLE OF INVENTION: And Compositions Based On No. 6083904ch Proteins And Nucleic Acids
TITLE OF INVENTION: Nucleic Acids
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/532.384
FILING DATE: 25-JUN-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/083,590
FILING DATE: 25-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-015
TELEPHONE: 212 790-9090
TELEFAX: 212 8698864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1078 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-532-384-11

Query Match 89.1%; Score 41; DB 3; Length 1078;
Best Local Similarity 90.0%; Pred. No. 5.5;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VCGSVLLSRK 10
Db 297 VCGSVLLSRK 306

RESULT 8
US-08-893-828-32
Sequence 32, Application US/08893828
Patent No. 6090922
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, Spyridon
APPLICANT: Muskavitch, Marc A.T.
APPLICANT: Fehon, Richard G.
APPLICANT: Rebay, Ilaria
APPLICANT: Blaumüller, Cristine M.
APPLICANT: Shepard, Scott B.
TITLE OF INVENTION: HUMAN NOTCH AND DELTA, BINDING DOMAINS
IN TOPOTHYMIC PROTEINS, AND METHODS BASED THEREON
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/893,828
FILING DATE: 11-JUL-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-050
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 1078 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-893-828-32

Query Match 89.1%; Score 41; DB 3; Length 1078;
Best Local Similarity 90.0%; Pred. No. 5.5;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VCGSVLLSRK 10
Db 297 VCGSVLLSRK 306

RESULT 9
US-08-185-432-17
Sequence 17, Application US/08185432
Patent No. 5750652
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, Spyridon
APPLICANT: Busseau, Isabelle
APPLICANT: Diederich, Robert J.
APPLICANT: Xu, Tian
APPLICANT: Matsuno, Kenji
TITLE OF INVENTION: DELTA PROTEINS, NUCLEIC ACIDS, AND ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/185,432
FILING DATE: 21-JAN-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-006
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 2556 amino acids

```
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-185-432-17

Query Match      89.1%; Score 41; DB 1; Length 2556;
Best Local Similarity 90.0%; Pred. NO. 14;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VCGSVLLSRK 10
Db 1751 VCGVLLSRK 1760

RESULT 10
US-08-083-590A-20
; Sequence 20, Application US/08083590A
; Patent No. 5786158
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, S. et al.
; TITLE OF INVENTION: Therapeutic And Diagnostic Methods
; TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And
; Nucleic Acids
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/083.590A
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 8698864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2556 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-083-590A-20

Query Match      89.1%; Score 41; DB 1; Length 2556;
Best Local Similarity 90.0%; Pred. NO. 14;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VCGSVLLSRK 10
Db 1751 VCGVLLSRK 1760

RESULT 11
US-08-532-384-20
; Sequence 20, Application US/08532384
; Patent No. 6083904
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, S. et al.
; TITLE OF INVENTION: Therapeutic And Diagnostic Methods
; TITLE OF INVENTION: And Compositions Based On No. 6083904ch Proteins And
; Nucleic Acids
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/532.384
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/083.590
; FILING DATE: 25-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 8698864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2556 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-532-384-20

Query Match      89.1%; Score 41; DB 3; Length 2556;
Best Local Similarity 90.0%; Pred. NO. 14;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VCGSVLLSRK 10
Db 1751 VCGVLLSRK 1760

RESULT 12
US-08-159-339A-979
; Sequence 979, Application US/08159339A
; Patent No. 6037135
; GENERAL INFORMATION:
; APPLICANT: Kubo, Ralph T.
; APPLICANT: Grey, Howard M.
; APPLICANT: Sette, Alessandro
; APPLICANT: Celis, Esteban
; TITLE OF INVENTION: HLA Binding peptides and Their
; NUMBER OF SEQUENCES: 1254
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
```

;; OPERATING SYSTEM: DOS
;; SOFTWARE: FastSEQ for Windows Version 2.0
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/159,339A
;; FILING DATE: 29-NOV-1993
;; CLASSIFICATION: 424
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/926,666
;; FILING DATE: 07-AUG-1992
;; APPLICATION NUMBER: US 08/027,746
;; FILING DATE: 05-MAR-1993
;; APPLICATION NUMBER: US 08/103,396
;; FILING DATE: 06-AUG-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Weber, Ellen Lauver
;; REGISTRATION NUMBER: 32,762
;; REFERENCE/DOCKET NUMBER: 018623-005030US
;; TELEPHONE: (415) 576-0200
;; TELEFAX: (415) 576-0300
;; TELEX:
;; INFORMATION FOR SEQ ID NO: 979:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 9 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; US-08-159-339A-979

Query Match 71.7%; Score 33; DB 3; Length 9;
Best Local Similarity 66.7%; Pred. No. 1.4e+05;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GSGVLLSRK 10
| | : | : | : |
Db 1 GGGLLMSRK 9

RESULT 13
US-08-159-339A-994
; Sequence 994, Application US/08159339A
; Patent No. 6037135
; GENERAL INFORMATION:
; APPLICANT: Kubo, Ralph T.
; APPLICANT: Grey, Howard M.
; APPLICANT: Sette, Alessandro
; APPLICANT: Celis, Esteban
; TITLE OF INVENTION: HLA Binding peptides and Their
; USES
; NUMBER OF SEQUENCES: 1254
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,339A
; FILING DATE: 29-NOV-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/926,666
; FILING DATE: 07-AUG-1992
; APPLICATION NUMBER: US 08/027,746
; FILING DATE: 05-MAR-1993

;; APPLICATION NUMBER: US 08/103,396
;; FILING DATE: 06-AUG-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Weber, Ellen Lauver
;; REGISTRATION NUMBER: 32,762
;; REFERENCE/DOCKET NUMBER: 018623-005030US
;; TELEPHONE: (415) 576-0200
;; TELEFAX: (415) 576-0300
;; TELEX:
;; INFORMATION FOR SEQ ID NO: 994:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 10 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; US-08-159-339A-994

Query Match 71.7%; Score 33; DB 3; Length 10;
Best Local Similarity 66.7%; Pred. No. 1.4;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GSGVLLSRK 10
| | : | : | : |
Db 2 GGGLLMSRK 10

RESULT 14
US-08-592-696-5
; Sequence 5, Application US/08592696
; Patent No. 5821334
; GENERAL INFORMATION:
; APPLICANT: Powers, Alvin C
; TITLE OF INVENTION: "INSULIN-DEPENDENT DIABETES
; MELLITUS-SPECIFIC CHIMERIC POLYPEPTIDES"
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Needle & Rosenberg, P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,696
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Selby Esq., Elizabeth
; REGISTRATION NUMBER: 38,298
; REFERENCE/DOCKET NUMBER: 22000.0043
; TELEPHONE: 404-688-0770
; TELEFAX: 404-688-9880
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 166 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-592-696-5

Query Match 71.7%; Score 33; DB 2; Length 166;
Best Local Similarity 66.7%; Pred. No. 27;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSGVLLSRK 10
| | : | : | : |
Db 99 GGGLLSMRK 107

RESULT 15

US-09-027-536-5
; Sequence 5, Application US/09027536
; Patent No. 5968757
; GENERAL INFORMATION:
; APPLICANT: Powers, Alvin C
; TITLE OF INVENTION: "INSULIN-DEPENDENT DIABETES
; TITLE OF INVENTION: MELLITUS-SPECIFIC CHIMERIC POLYPEPTIDES"
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Needle & Rosenberg, P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/027,536
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/592,696
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Selby Esq., Elizabeth
; REGISTRATION NUMBER: 38,298
; REFERENCE/DOCKET NUMBER: 22000.0043
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-688-0770
; TELEFAX: 404-688-9880
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 166 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-027-536-5

Query Match 71.7%; Score 33; DB 2; Length 166;
Best Local Similarity 66.7%; Pred. No. 27;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSGVLLSRK 10
| | : | : | : |
Db 99 GGGLLSMRK 107

Search completed: September 6, 2001, 16:39:40
Job time: 129 sec

